DOCUMENT RESUME

ED 062 995	LI 003 676
TITLE	Proceedings of the International Symposium: "UDC in Relation to Other Indexing Languages" (Herceg Novi, Yugoslavia, June 28-July 1, 1971).
INSTITUTION	International Federation for Documentation, The Hague (Netherlands).; Yugoslav Center for Technical and Scientific Documentation, Belgrade.
PUB DATE	71
NOTE	417p.; (156 References)
AVAILABLE FROM	Jugoslovenski Centar za Tehnicu i Naucnu Dokumentaciju, P. O. Box 724, Belgrade, Yugoslavia ($6.00)
EDRS PRICE	MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Classification; *Indexing; Information Networks; Information Retrieval; *Information Systems; *Interdisciplinary Approach; International Programs; *Language Standardization; *Multilingualism; Seminars; Subject Index Terms; Symposia
IDENTIFIERS	Multidisciplinary Information Systems; Multilingual Information Systems; UDC; *Universal Decimal Classification

ABSTRACT

The stress of this symposium was on the role of the Universal Decimal Classification (UDC) and other indexing languages, especially in multilingual and multidisciplinary information systems, rather than on its potential uses in mechanized retrieval systems. The consensus of the Symposium was: (1) that a system like the UDC cannot, without major structural changes, continue to serve as an internationally accepted ordering system for the rapidly proliferating new subject-fields that have to be handled in modern information retrieval systems; (2) UDC improvement should be based on better methodology and clearer guidelines for those working to revise specific fields and (3) that the International Federation for Documentation (FID) should aim at preparing, publishing and maintaining a Standard Reference (SR) scheme consisting of a "roof classification schedule" of not more than three or four hierarchic levels as the unifying/switching language, with suitably indicated points and symbol devices for coupling to specific classifications, thesauri, etc. leaving the responsibility for full editions with national members and special-subject editions with the specialized interests concerned. A list of symposium participants is appended. (Author/NH)

ED 062995

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

L. Dulovic

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

(2)

PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM:

"UDC in Relation to other Indexing Languages"

~~Held in~~

(HERCEG NOVI, Yugoslavia, June 28 — July 1, 1971).

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

(3) Yugoslav Center for Technical and Scientific Documentation,
~~11.001~~ Beograd, ~~P. O. B. 724~~, Yugoslavia.
and
International Federation for Documentation,
(4) 7. ~~7 Hofweg~~, The Hague, (Netherlands).

3BB03547

LI 003 676

422 p.
156 Ref.

SYMPOSIUM STEERING COMMITTEE

1. Mr.Ljubomir Dulović
2. Mr.Milutin Gavrilov
3. Mrs.Bogomila Živković
4. Mr.Djordje Nikolić
5. Mr.Bogdan Jerković
6. Mrs.Jana Borovnjak

CONTENTS

1. Editorial.. 1 pp.

2. Foreword... 1 pp.

3. Cooke,G.A., Heaps,D., Mercier,M.:
   THE STUDY OF UDC AND OTHER INDEXING LANGUAGES THROUGH
   COMPUTER MANIPULATION OF MACHINE READABLE DATA BASES.... 1-24 pp.

4. Cruzeiro,Maria M.:
   INDEXING LANGUAGES ON ECONOMICS.UDC,UNESCO AND OECD..... 1-6 pp.

5. Černyj,A.I.:
   THE LOGICAL PRINCIPLES OF THE DEVELOPMENT OF UDC AND
   THE INFORMATION RETRIEVAL LANGUAGES OF DESCRIPTOR TYPE
   (in Russian).. 1-28 pp.

6. Dahlberg,I.:
   THE UDC AND AN IDEAL INDEXING LANGUAGE................. 1-25 pp.

7. Dulović,Lj., Gavrilov,M.:
   PROBLEMS OF INDEXING AND CLASSIFICATION IN SMALL,
   DEVELOPING AND MULTILINGUAL COUNTRIES.................. 1-6 pp.

8. Hindson,R.:
   INDEXING AND RETRIEVAL OF RECORDS OF BRITISH PATENT
   SPECIFICATIONS USING THE UNIVERSAL DECIMAL CLASSIFICATION AND THE BISRA EXPERIMENTAL THESAURUS............. 1-52 pp.

9. Kara (Marosi),A.:
   THE RELATION OF THE UDC TO THE EURATOM THESAURUS FROM
   THE POINT OF VIEW OF THE SMALL INFORMATION CENTRE....... 1-15 pp.

lo.Kofnovec,L.,Simandl,D.:
   OPTIMIZING THE STRUCTURE OF CLASSIFICATION AND
   INDEXING SYSTEMS.. 1-19 pp.

11.Mole,M.:
   UDC - INSTRUMENT FOR CONNECTING SCIENTIFIC INFORMATION,
   EDUCATION AND PRACTICE.................................. 1-2 pp.

12. Neelameghan,A.:
A THEORETICAL FOUNDATION FOR UDC: ITS NEED AND
FORMULATION.. 1-52 pp.

13. Öhman,E., Olivecrona,Ch.:
SOME NOTATIONAL, HIERARCHIC, AND SYNTACTIC PROBLEMS IN
CONNECTION WITH CONCORDENCES BETWEEN UDC AND THESAURI... 1-48 pp.

14. Perreault,J.M.:
THE CONCEPT "MAIN CLASS" AND ITS IMPLICATION FOR
STRUCTURAL IMPROVEMENT OF UDC.......................... 1-21 pp.

15. Rigby,M.:
AN INVERTED UDC INDEX AS A STEP TOWARD SIMPLIFIED AND
EFFECTIVE THESAURUS DEVELOPMENT AND MAINTENANCE......... 1-8 pp.

16. Samuelson,K.:
RELAY-SWITCHES AND REFERRAL - DIRECTORIES FOR
INTERNATIONAL DATA-BASES AND INFORMATION NETWORKS....... 1-12 pp.

17. Ścibor,E.:
SOME REMARKS ON THE COMPARATIVE EFFICIENCY OF UDC AND
DESCRIPTOR LANGUAGES................................... 1-10 pp.

18. Ungurian,O.:
THE POSSIBILITY TO USE THE UDC AS A CO-ORDINATE INDEXING
LANGUAGE (in Russian).................................. 1-9 pp.

19. Wåhlin,E.:
THE AR COMPLEX.ADAPTED SYSTEMS USED TOGETHER WITH A
COMMON REFERENCE SYSTEM................................ 1-24 pp.

20. Wellisch,H.:
A CONCORDANCE BETWEEN UDC AND TEST (THESAURUS OF
ENGINEERING AND SCIENTIFIC TERMS): RESULTS OF A
PILOT PROJECT.. 1-33 pp.

21. De Wijn,J.H.:
THE UDC AS A LANGUAGE OF INFORMATION ITSELF............ 1-12 pp.

22. Lloyd,G.A.:
SUMMARY OF SYMPOSIUM CONSENSUS AND POSSIBLE FUTURE
TRENDS... 1-3 pp.

23. Concluding speech of Mr. Ljubomir Dulović............. 1 pp.
24. List of participants.................................... 1-4 pp.

EDITORIAL

The Symposium reported herein was initiated at FID Second Seminar on UDC and Mechanized Information Systems (Frankfurt/M, June 1-5, 1970) by Ljubomir Dulović, Director, Yugoslav Center for Technical and Scientific Documentation, Belgrade. There were a few objectives, among which the most important were to confront different views and opinions and to come to a conclusion.

The Symposium attended 120 participants (26 from 13 foreign countries and 94 from Yugoslavia).

In publishing proceedings which include 19 papers, the organizers express their gratitude to all those who cooperated and contributed to its realization.

Yugoslav Center for Technical and
Scientific Documentation

FOREWORD

This 'Herceg Novi Symposium', held in a pleasant Yugoslav Adriatic resort, was as international as the two preceding seminars in Copenhagen (September 1968) and Frankfurt (June 1970) but differed in that the stress here was on the role of UDC and other indexing languages, especially in multilingual and multidisciplinary information systems, rather than on its potential uses in mechanized retrieval systems. This was a particularly appropriate emphasis, since the FID's National member, Jugoslovenski centar za tehničku i naučnu dokumentaciju, who organized the Symposium under the joint sponsorship of three FID committees (FID/CCC, FID/CR and FID/TM), itself faces the difficult information exchange problems of a multilingual federal state, in which three separate UDC Abridged editions have already been published in Serbo-Croat, Slovene and Macedonian.

Important as this aspect was, however, and still topical though the mechanization problems of the UDC remained, it may well be that this Symposium will be remembered mainly as a point of no return in the thinking and conclusions about the UDC and its structural improvement within the framework of the CCC's development programme for the 1970s. A glance at the "Summary of consensus and possible future trends", in which an attempt is made to draw together the views of my fellow panel membres and those expressed by many of the contributors of papers and other participants in the discussions, will show that the efforts of the FID and CCC will have to be concentrated - without abandonment of the normal routine revision procedure - on the design of a new superstructure to match the needs of UNISIST and similar major information networks, without destroying the many sound elements of the existing UDC. This trend was given impetus at a meeting of the recently established UDC Advisory Panel (Chairman Dr. H. Coblans), whose recommendation to establish a new Standard Reference Code (SRC) derived from the UDC was adopted by a meeting of the full CCC in Warsaw, 26-29 October 1971.

It is to be hoped that the Warsaw decisions - now widely circulating - will be welcomed by most existing UDC users (despite some unavoidable inconveniences they may have to face) and by many new and potential users, and that the plans will be implemented sufficiently during the next 3-4 years to demonstrate the suitability of the new SRC as an international unifying /switching language for projekcts like UNISIST, as well as for detailed documentation, indexing and retrieval purposes which the UDC has served, with relatively minor modification, for so many years. If the implementation of these decisions and plans comes to fruition and the SRC gains wide acceptance, it will be thanks in no small measure to the impetus of this Symposium and the initiative of its Yugoslav organizers.

31 January, 1972

G.A. Lloyd, Head
FID Classification
Department,
Secretary to FID/CCC

THE STUDY
OF UDC AND OTHER INDEXING LANGUAGES THROUGH COMPUTER MANIPULATION OF MACHINE-READABLE DATA BASES

Geraldine A.COOKE,Doreen M.HEAPS,Marcel MERCIER
University of Alberta,Edmonton,Alberta,Canada

INTRODUCTION

Information Identification and Indexing Languages

The principal problem in the handling of information-bearing records is the correct identification of the content of the record. In addition, at some point, each record must be stored and its storage location identified.

Correct content or subject identification is very difficult. The identifier must have specialist knowledge of the subject, must make subjective judgements concerning the values of the different themes in the record and must fit these into an overall conceptual scheme which exists in both time and place (1). It must also be remembered that this conceptual scheme will differ at different times and in different places.

The procedure of identifying content or subject is called indexing or classification. All such classification or similar subject identification methods may be regarded as the use of specific indexing languages.

Furthermore, each indexing language may also be regarded as an information language with its own vocabulary and syntax. For example, we may consider index words as the vocabulary of an information language and the indexing rules as the syntax, or we may consider a classification notation the metalanguage of an information language. Then the symbols of the notation comprise the vocabulary and the rules for the combination of the symbols make up the syntax. Freeman, Gardin, and others have discussed the concept of an information language at some length in recent papers (2) (3)(4).

This paper outlines a method for the comparison and evaluation of indexing languages that employs computer techniques and machine-readable data bases. The techniques may also be of assistance in the formal analysis of the UDC and other indexing languages when these are regarded as information languages.

Indexing Language Requirements

If we regard subject assignment as the main problem in the identification of records, then the proper use and the correct choice of a suitable indexing language is fundamental to the efficient handling of information. A suitable indexing language will help the indexer to achieve consistency, efficiency, and responsiveness.

Much effort has gone into ensuring indexer consistency in the use of indexing languages. There are rules for the building of thesauri and for indexing techniques (5) (6). The United States Library of Congress trains its classifiers and issues authority lists (7). Similar aids are available to the UDC classifier. Despite such aids, consistency can usually only be achieved through detailed and tedious manual checking.

Much attention has also been paid to judging and testing indexing language efficiency, especially when it is related to user satisfaction. The Cranfield studies and the work of Lancaster are prime examples of investigations of indexing language efficiency (8) (9).

Far less attention has been paid, however, to the requirement for responsiveness to change or for appropriateness in varying situations. There have been few studies reported in the open literature on the relationships, such as similarities and differences, between various indexing languages (10), or on the linking of languages to varying user groups (11), or on the interaction of subject words or index terms with the classification notation (12), or on the changes in indexing languages over time, in a situation where it is necessary for index terms to reflect current usage. The lack of investigation in these areas is understandable because such studies require much time and effort in word counting and comparisons that, on the whole, up to the present, would have had to be done manually.

Computer Assistance

The computer and large machine-readable data bases allow us, now, to match vocabulary against vocabulary, subject list against subject list, classification notation against classification notation and each one against any other. We may be able to discover if a subject field which is described by a certain vocabulary can be appropriately indexed by a specific indexing language, especially if we have a classification notation that can be regarded as the metalanguage for that indexing language (13). We may do this, for example, by checking the notation against the general area covered by the subject words. Let us suppose that a word with related terms occurs in a thesaurus. If we have available a MARC tape or similar classified data base, we can run the thesaurus word against the subject words on the tape to ascertain what classification occurs on the tape. Then we can check the thesaurus cluster to see if it matches the class. In other words, through the use of available machine-readable data bases and appropriate programs we can automate the type of steps that are taken manually in using a tool such as the English Electric Thesaurofacet (14).

It is clear that the modern computer facility with on-line and batch capability coupled with the availability of machine-readable data bases provides a tool whereby indexing languages can be studied, controlled, and developed with increased precision.

DATA BASE MANIPULATION

In the Department of Computing Science, University of Alberta, beginning in the summer of 1970, machine-readable data bases have been used in investigations of class and term or indexing language relationships. The investigation reported in this paper is part of a major project concerned with the development of an on-line thesaurus-based information storage and retrieval system for water resources management. In this system indexing and searching will normally both be done at the terminal through access to an on-line thesaurus, which is linked to a data base, somewhat MARC-like in format, but which carries UDC numbers in one of the fields. A batch version of the system has also been developed.

In planning this system a question arose regarding the choice of an appropriate indexing language. The indexing language whose vocabulary will make up the on-line thesaurus is the key to the entire system. As a preliminary to the choice of a language it was decided to investigate and compare the appropriateness of the UDC, the LC, and the United States Water Resources Thesaurus as indexing languages for the control of water resources planning literature. This is being carried out through the manipulation of various machine-readable data bases as described below.

Data Bases

The data bases used for this project are the MARC tapes, the UDC English Language Master Files, the United States Water Resources Scientific Information Center Thesaurus (WRSIC), Geology Document File, EMR Word List, and coded documents from the Information System of the Planning Division of the Canada Department of Forestry and Fisheries;

the Division was formerly part of the Canada Department of Energy, Mines, and Resources (EMR).

MARC Tapes. The MARC tapes (Machine-Readable Catalog) consist of magnetic tapes distributed by the Library of Congress every two weeks on subscription. There are approximately 1000 titles per issue; they cover current monographs in all subject areas. Each entry is the image of its corresponding catalog card. The MARC tapes issued during 1970 were made available for use by the University of Alberta Library Systems Group.

UDC English Language Master Files. This file, on magnetic tape, includes the complete merged set of UDC schedules accumulated by the AIP/UDC project from 1965 to 1967, a total of 110,759 records (15). Each record includes the UDC number, a code which represents the source of the record and the English language equivalent of the UDC number. Many records also contain cross references and scope notes, which serve to define and delimit a concept further.

The sources of input to the UDC English Language Master File and their codes are as follows:

ENAB61	Abridged Edition, UDC, 1961 complete
ENED65	Education Edition, 1965
ENFU--	Full Edition, unpublished ms.
ENFU43	Full Edition, 1943 (partial)
ENFU54	Full Edition, 1954 (partial)
ENFU55	Full Edition, 1955 (partial)
ENFU58	Full Edition, 1958 (partial)
ENFU64	Full Edition, 1964 (partial)
ENFU67	Medium Edition, 1967 (not yet published)

(EN is the language code meaning English)

These tapes were made available by Robert Freeman, who took part in the AIP/UDC project.

Water Resources Thesaurus (WRSIC). This thesaurus, developed by the Water Resources Scientific Information Center, now part of the United States Department of the

Interior, consists of 4039 main terms and their cross referenc interfiled with 1141 other terms (use references), that are intended to lead the user to the proper descriptors. A tape-copy of this thesaurus was provided by the Computer Center at the University of Manitoba. The tape consists of an exact copy of the main body of the thesaurus; this was issued in November 1966.

EMR Wordlist. A list of words also incorporated in this study was supplied by the Canada Department of Fisheries and Forestry. This list is to supplement the Water Resources Thesaurus. These new words will, in effect, Canadianize the thesaurus; the majority of the terms are those that are relevant to Canadian policy formation and management of water resources.

Geology Document File. This is a tape-copy of a file of punched cards containing title and UDC numbers used in compiling the UNIDEK Index to GeoScience Abstracts (a part of the AIP/UDC project). There are approximately 7300 document entries. This was supplied by Robert Freeman.

Planning Division Data Base. This file consists of approximately 400 documents that have been keyworded, abstracted, and tentatively coded by the Resource Data Section of the Department of Forestry and Fisheries. The material covered consists of research reports, monographs, statutes, and some applications and awards for research grants.

General Methodology

In order to investigate the suitability of the LC and the UDC for the classification of water resources information, two different approaches have been taken.

In the first approach the UDC was studied by the manipulation of the UDC file of schedules in conjunction with the Water Resources Thesaurus. The Geology Document File and the Planning Division Data Base are also used here. In the second approach the LC is studied through a somewhat similar manipulation of the MARC tapes and the Water Resources Thesaurus.

It is necessary to use two approaches because of the differences that occur in the two classification schemes and because of the limited availability of mechanized data bases with which to work. No mechanized schedules exist for the LC;whereas the mechanized UDC schedules, which resulted from the AIP/UDC project, form the basis of this UDC investigation. This problem has been partially circumvented through the use of the LC subject headings and classifications on the MARC tapes. The MARC tapes, rather than the LC subject heading tape,have been used for several reasons, one of them being that the MARC tapes represent the only LC classified data base regularly available.

The procedure employed in both approaches is further described below.

Investigation of the UDC

The work underway in investigating the UDC as a tool for the classification of water resources material will be described in three steps. 1) The first step finds all subject headings in the UDC that are similar to the descriptors from the Water Resources Thesaurus. 2) The second step analyzes the subject headings extracted in the first step. These two steps result in a concordance between the Water Resources Thesaurus and the UDC schedules; this concordance is used in step three. 3) Step three determines the suitability of using the UDC in conjunction with the Water Resources Thesaurus for the classification of information for water resource management. This step includes user studies.

Step 1. This step is confined to the manipulation of the machine-readable UDC schedules and the WRSIC thesaurus (See

Figure 1). Both files are converted into a form suitable for automatic comparison purposes. This means that the UDC schedules are stripped of all codes and non-significant words such as the, and, which, etc. This operation results in a file that consists of all subject classes containing the significant words used in their descriptions and the corresponding UDC number.

The Water Resources Thesaurus is then partially stripped of cross references to provide a file, with no duplications, of all terms used in the thesaurus. In effect, the resulting file contains only main terms and use references. This file is then run through other programs to give it a format suitable for comparison with the reformatted UDC file.

Because of the new form of the UDC file and of the thesaurus file all UDC subject classes found contain not only the exact matches with the descriptor terms, but also different formats of the terms. An example of the match procedure is shown in Figure 2. In this example the thesaurus term "drinking water" would be found matched to any UDC subject classes containing the English language equivalent in any of the following forms:

-drinking water
-drinking of the water
-water for drinking
-water purification, drinking, etc.

In fact, this term was found in four subject classes (628.1, 644.61, 663.6 and 696.11). The method used for finding a descriptor term in any of these forms leads to some non-relevant classes but these are few and obvious. This technique is very useful for finding terms such as "elastic theory" in class 539.31 -Elasticity. Elastic Forces. General Theory of Stress.

At present it cannot be said which specific classes of the schedules stand out more than other classes when compared

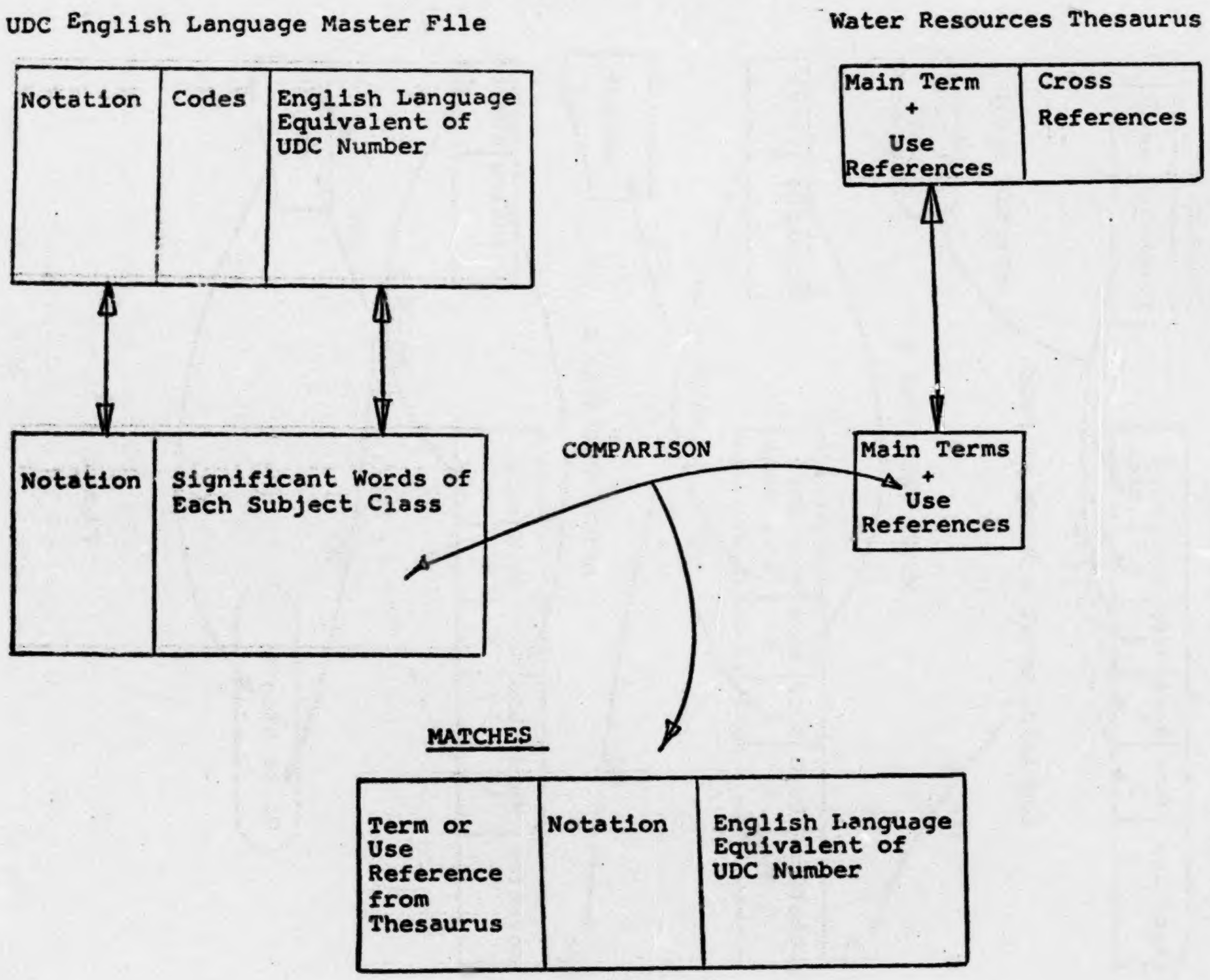

UDC English Language Master File
Notation
Codes
English Language Equivalent of UDC Number
Notation
Significant Words of Each Subject Class
Water Resources Thesaurus
Main Term + Use References
Cross References
Main Terms + Use References
COMPARISON
MATCHES
Term or Use Reference from Thesaurus
Notation
English Language Equivalent of UDC Number

Figure 1

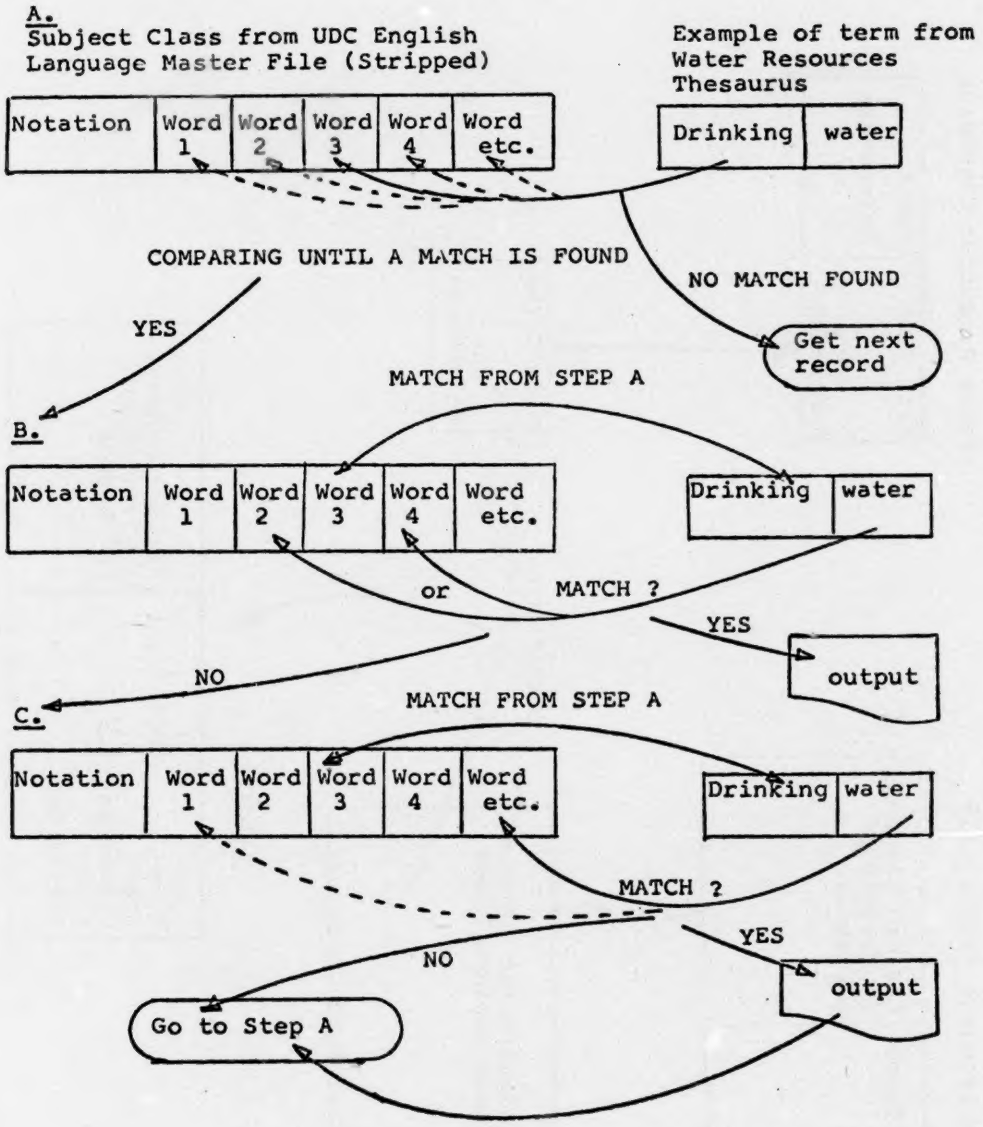

A.
Subject Class from UDC English Language Master File (Stripped)
Example of term from Water Resources Thesaurus
Notation
Word 1
Word 2
Word 3
Word 4
Word etc.
Drinking
water
COMPARING UNTIL A MATCH IS FOUND
NO MATCH FOUND
Get next record
YES
MATCH FROM STEP A
B.
Notation
Word 1
Word 2
Word 3
Word 4
Word etc.
Drinking
water
or
MATCH ?
YES
output
NO
C.
MATCH FROM STEP A
Notation
Word 1
Word 2
Word 3
Word 4
Word etc.
Drinking
water
MATCH ?
YES
NO
output
Go to Step A

Figure 2

with the thesaurus, or what classes are not very relevant, as the analysis of the resulting files has not been completed. However, sixty per cent of the matches occur in classes 547/699. Ten per cent of the matches occur in class 3. About ten per cent in class 7, and the rest, approximately twenty per cent, occur in the remaining schedules including the auxiliaries. As the file now stands, the number of matches occurring seems to be directly proportional to the number of subject classes in each of the main classes of the schedules. This proportionate spread of terms seems to indicate that a subset of the schedules would not be adequate for classifying material in the field of water resources, but that the overall set would be needed. However, it already appears that some classes will not need to be as detailed as others. A set of schedules for water resources might be similar in form to those used for other special collections such as the Scott Polar Library (16).

Of the 5200 words (that is, words from the thesaurus) that were tested against the UDC schedules there are approximately 54% in which no match occurs. It is estimated that this number will become smaller once the matched list is studied further and the keypunched up-dates to schedules 551 and 556 are incorporated in the study. As stated, this list has not yet been completely analyzed, but it is suspected that some of the terms did not produce matches because of differences in spelling, or because they consisted of concepts that could not be covered by only one subject class and that in UDC collections would be commonly coloned.

Step 2. Step 2 consists of analyzing the file resulting from Step 1. This file, comprising approximately 49,000 matches of subject classes with thesaurus terms, is sorted automatically according to UDC number. The file is then sorted by matched thesaurus terms to produce a second file with the thesaurus terms in alphabetical order.

Both these files are then printed as List 1 and List 2 respectively. Each printout consists of the word that was found, the UDC number for the subject class it was found in, and the English language equivalent for that class. Samples from these two lists are shown in Table 1.

These lists are now being used to "link" thesaurus terms to subject classes. Since many of the terms occur only a few times it is not difficult to see, by using List 2 only, if the matched class would serve as a relevant 'entry point' or 'link' for that term. It is more difficult to identify entry points for terms that occur frequently. Many terms, such as air, acids, buildings, match a thousand times or more. Such a large number of matches is generally caused by a term, such as acids, appearing in a 'roof class' and in many of its subclasses as well.

As yet, no set procedure for the linking of the thesaurus terms to the subject classes has emerged because each matched term must be evaluated separately. In spite of difficulties, certain basic steps that are been used for most terms have been established. The following is a description of these:

Take a unique term in List 2. For example, from the subset of List 2 illustrated in Table 1(b) the first term to be looked at would be "cements". Then the thesaurus would be checked to see the relationships of the term in the thesaurus (Table 1(c)). In the on-line implementation the relationship would be displayed automatically by accessing the thesaurus. Then a check is performed of the order of the UDC numbers for the classes in which this word occurs to see if some type of structure is evident. From the structure a possible 'entry point' or 'link' to the schedules may emerge. From List 2 in Table 1 the possible major entry points for

TABLE 1

A

List 1

Thesaurus Entry	Notation	English Language Equivalent
PORTLAND CEMENTS	666.94	CEMENTS
CEMENTS	666.94	CEMENTS
PORTLAND CEMENTS	666.942	PORTLAND CEMENTS
CEMENTS	666.942	PORTLAND CEMENTS
CEMENTS	666.948	ALUMINOUS CEMENTS
MATERIALS	666.95	SULFATE, TRASS CEMENTS
CEMENTS	666.95	SULFATE, TRASS SEMENTS
MATERIALS	666.96	HARDSETTING CEMENT COMPOSITIONS
MORTAR	666.97	CONCRETE
UNDERWATER	666.97.0.15	UNDERWATER SETTING
TREATMENT	666.97.0.3	PREPARATION. PROCESSING. TREATMENT
MACHINERY	666.97.0.5	PLANT, MACHINERY AND EQUIPMENT

B

List 2

Thesaurus Entry	Notation	English Language Equivalent
CEMENTS	666.94	CEMENTS
CEMENTS	666.942	PORTLAND CEMENT
CEMENTS	666.948	ALUMINOUS CEMENTS
CEMENTS	666.95	SULFATE, TRASS CEMENTS
CEMENTS	691.5	BINDING AND BEDDING MATERIALS
CEMENTS	691.54	CEMENTS
CEMENTS	691.541	ROMAN CEMENTS
CEMENTS	691.544	OTHER CEMENTS
CENSUS	312	DEMOGRAPHY. POPULATION. STATIS...
CENSUS	351.755.3	CENSUS REGISTERS. C.F. 312
CENSUS	656.0.21	NUMBER AND FREQUENCY OF TRANSPORT IN GENERAL. TRAFFIC SURVEYS.
CENTRIFUGAL PUMPS	621.67	CENTRIFUGAL PUMPS. AXIAL FLOW AND TURBO-PUMPS
CENTRIFUGAL PUMPS	621.671	CENTRIFUGAL PUMPS IN GENERAL
CENTRIFUGAL PUMPS	621.671.1	OPEN CENTRIFUGAL PUMPS
CENTRIFUGAL PUMPS	621.671.2	CLOSED CENTRIFUGAL PUMPS
CENTRIFUGAL PUMPS	621.671.5	CENTRIFUGAL PUMPS WITH INVOLUTE BLADES

C

Thesaurus Form

Cements
  NT Portland Cements
  RT Adhesives
    -Aggregates
     Alkali-Aggregate Reaction
     Asphalt
     Cement Grouting
    -Clays
    -Concrete Additives (etc. as in Table 2)
     Concrete Mixes

the word "cements" are 666.9, since four of the matches are subsets of this class, and 691.5, again because four of the matches are subsets of this class. Once possible entry points have been determined a check is made of the English language equivalent for that class to see if it is relevant or not. Sometimes List 1, sorted by UDC number, and/or the schedules must be consulted in order to help determine relevancy.

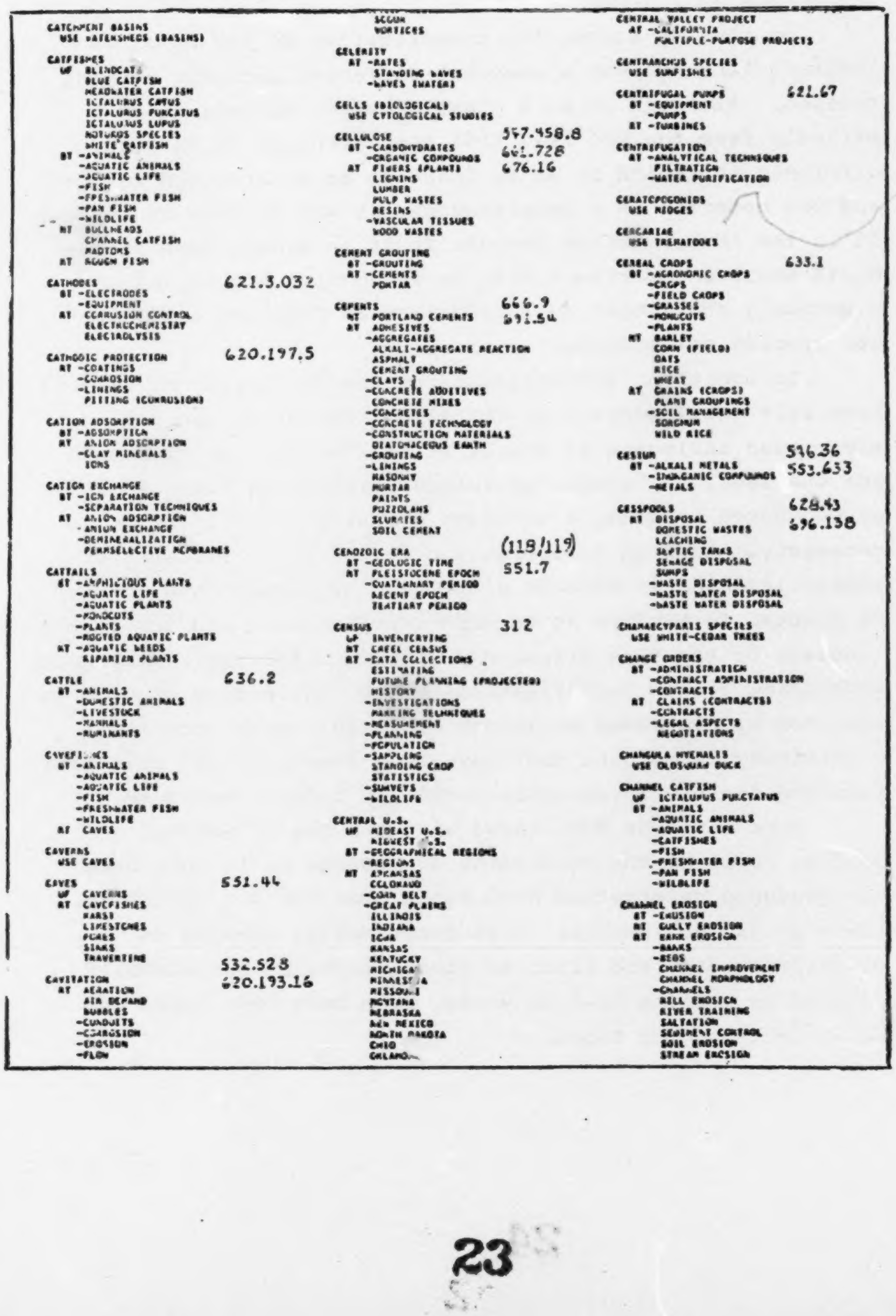
Once all non-relevant entry points are eliminated it is necessary to decide on the validity of those left. In this particular use this is done by keeping those classes which specifically serve the classification of water resources information and discarding those that do not. It must also be decided whether, for the use envisaged, the entry point is too high or too low in the structure of the class being entered.

With this procedure entry points for the term "cements" would be 666.9 and 691.5; entry points for "census" would be 312; entry point for "centrifugal pumps" would be 621.671. For "centrifugal pumps" the entry would be 621.671 because an entry at 621.67 covers axial-flow and turbo-pumps also. It should be realized that List 1 in Table 1 is a subset of the complete list that would be used in making the decisions discussed above.

Step 3. The concordance of the Water Resources Thesaurus with the UDC resulting from the previous two steps is to be tested against the Planning Division Data Base and the Geology Document File. This modified thesaurus is to be used both as a guide to indexing documents, and as a guide to the automatic retrieval of documents via UDC number and/or keywords. An example of the tentative form of the final concordance is given in Table 2.

User tests of the system will then be carried out. The test users will be those involved in the planning and policy

TABLE 2



```
CATCHMENT BASINS
  USE WATERSHEDS (BASINS)

CATFISHES
  UF   BLINDCATS
       BLUE CATFISH
       HEADWATER CATFISH
       ICTALURUS CATUS
       ICTALURUS FURCATUS
       ICTALURUS LUPUS
       NOTURUS SPECIES
       WHITE CATFISH
  BT -ANIMALS
      -AQUATIC ANIMALS
      -AQUATIC LIFE
      -FISH
      -FRESHWATER FISH
      -PAN FISH
      -WILDLIFE
  NT   BULLHEADS
       CHANNEL CATFISH
       MADTOMS
  RT   ROUGH FISH

CATHODES                                621.3.032
  BT -ELECTRODES
      -EQUIPMENT
  RT   CORROSION CONTROL
       ELECTROCHEMISTRY
       ELECTROLYSIS

CATHODIC PROTECTION                     620.197.5
  RT -COATINGS
      -CORROSION
      -LININGS
       PITTING (CORROSION)

CATION ADSORPTION
  BT -ADSORPTION
  RT   ANION ADSORPTION
      -CLAY MINERALS
       IONS

CATION EXCHANGE
  BT -ION EXCHANGE
      -SEPARATION TECHNIQUES
  RT   ANION ADSORPTION
      -ANION EXCHANGE
      -DEMINERALIZATION
       PERMSELECTIVE MEMBRANES

CATTAILS
  BT -AMPHIBIOUS PLANTS
      -AQUATIC LIFE
      -AQUATIC PLANTS
      -MONOCOTS
      -PLANTS
      -ROOTED AQUATIC PLANTS
  RT -AQUATIC WEEDS
       RIPARIAN PLANTS

CATTLE                                  636.2
  BT -ANIMALS
      -DOMESTIC ANIMALS
      -LIVESTOCK
      -MAMMALS
      -RUMINANTS

CAVEFISHES
  BT -ANIMALS
      -AQUATIC ANIMALS
      -AQUATIC LIFE
      -FISH
      -FRESHWATER FISH
      -WILDLIFE
  RT   CAVES

CAVERNS
  USE CAVES

CAVES                                   551.44
  UF   CAVERNS
  RT   CAVEFISHES
       KARST
       LIMESTONES
       POOLS
       SINKS
       TRAVERTINE
                                        532.528
CAVITATION                              620.193.16
  RT   AERATION
       AIR DEMAND
       BUBBLES
      -CONDUITS
      -CORROSION
      -EROSION
      -FLOW
```

```
       SCOUR
       VORTICES

CELERITY
  RT -RATES
       STANDING WAVES
      -WAVES (WATER)

CELLS (BIOLOGICAL)
  USE CYTOLOGICAL STUDIES

CELLULOSE                               547.458.8
  BT -CARBOHYDRATES                     661.728
      -ORGANIC COMPOUNDS                676.16
  RT   FIBERS (PLANT)
       LIGNINS
       LUMBER
       PULP WASTES
       RESINS
      -VASCULAR TISSUES
       WOOD WASTES

CEMENT GROUTING
  BT -GROUTING
  RT -CEMENTS
       MORTAR

CEMENTS                                 666.9
  NT   PORTLAND CEMENTS                 691.54
  RT   ADHESIVES
      -AGGREGATES
       ALKALI-AGGREGATE REACTION
       ASPHALT
       CEMENT GROUTING
      -CLAYS
      -CONCRETE ADDITIVES
       CONCRETE MIXES
      -CONCRETES
      -CONCRETE TECHNOLOGY
      -CONSTRUCTION MATERIALS
       DIATOMACEOUS EARTH
      -GROUTING
      -LININGS
       MASONRY
       MORTAR
       PAINTS
       POZZOLANS
       SLURRIES
       SOIL CEMENT
                                        (118/119)
CENOZOIC ERA
  BT -GEOLOGIC TIME                     551.7
  NT   PLEISTOCENE EPOCH
      -QUATERNARY PERIOD
       RECENT EPOCH
       TERTIARY PERIOD

CENSUS                                  312
  UF   INVENTORYING
  NT   CREEL CENSUS
  RT -DATA COLLECTIONS
       ESTIMATING
       FUTURE PLANNING (PROJECTED)
       HISTORY
      -INVESTIGATIONS
       MARKING TECHNIQUES
      -MEASUREMENT
      -PLANNING
      -POPULATION
      -SAMPLING
       STANDING CROP
       STATISTICS
      -SURVEYS
      -WILDLIFE

CENTRAL U.S.
  UF   MIDEAST U.S.
       MIDWEST U.S.
  BT -GEOGRAPHICAL REGIONS
      -REGIONS
  NT   ARKANSAS
       COLORADO
      -CORN BELT
      -GREAT PLAINS
       ILLINOIS
       INDIANA
       IOWA
       KANSAS
       KENTUCKY
       MICHIGAN
       MINNESOTA
       MISSOURI
       MONTANA
       NEBRASKA
       NEW MEXICO
       NORTH DAKOTA
       OHIO
       OKLAHO...
```

```
CENTRAL VALLEY PROJECT
  RT -CALIFORNIA
       MULTIPLE-PURPOSE PROJECTS

CENTRARCHUS SPECIES
  USE SUNFISHES

CENTRIFUGAL PUMPS                       621.67
  BT -EQUIPMENT
      -PUMPS
  RT -TURBINES

CENTRIFUGATION
  RT -ANALYTICAL TECHNIQUES
       FILTRATION
      -WATER PURIFICATION

CERATOPOGONIDS
  USE MIDGES

CERCARIAE
  USE TREMATODES

CEREAL CROPS                            633.1
  BT -AGRONOMIC CROPS
      -CROPS
      -FIELD CROPS
      -GRASSES
      -MONOCOTS
      -PLANTS
  NT   BARLEY
       CORN (FIELD)
       OATS
       RICE
       WHEAT
  RT   GRAINS (CROPS)
       PLANT GROUPINGS
      -SOIL MANAGEMENT
       SORGHUM
       WILD RICE
                                        546.36
CESIUM
  BT -ALKALI METALS                     553.633
      -INORGANIC COMPOUNDS
      -METALS
                                        628.43
CESSPOOLS
  RT   DISPOSAL                         696.138
       DOMESTIC WASTES
       LEACHING
       SEPTIC TANKS
       SEWAGE DISPOSAL
       SUMPS
      -WASTE DISPOSAL
      -WASTE WATER DISPOSAL
      -WASTE WATER DISPOSAL

CHAMAECYPARIS SPECIES
  USE WHITE-CEDAR TREES

CHANGE ORDERS
  BT -ADMINISTRATION
      -CONTRACT ADMINISTRATION
      -CONTRACTS
  RT   CLAIMS (CONTRACTS)
      -CONTRACTS
      -LEGAL ASPECTS
       NEGOTIATIONS

CHANGULA HYEMALIS
  USE OLDSQUAW DUCK

CHANNEL CATFISH
  UF   ICTALURUS PUNCTATUS
  BT -ANIMALS
      -AQUATIC ANIMALS
      -AQUATIC LIFE
      -CATFISHES
      -FISH
      -FRESHWATER FISH
      -PAN FISH
      -WILDLIFE

CHANNEL EROSION
  BT -EROSION
  NT   GULLY EROSION
  RT   BANK EROSION
      -BANKS
      -BEDS
       CHANNEL IMPROVEMENT
       CHANNEL MORPHOLOGY
      -CHANNELS
       RILL EROSION
       RIVER TRAINING
       SALTATION
       SEDIMENT CONTROL
       SOIL EROSION
       STREAM EROSION
```

formation of water resources management in Canada.

Investigation of the LC

As already stated, the investigation of the LC as an indexing language has a somewhat different form, for varying reasons. Although LC as a classification differs quite markedly from the UDC (17) (18) and, although it is more difficult to regard it as an indexing or information language and the notation as a metalanguage, it was decided to include it in the investigation because it is so widely used on the North American continent and, in the United States, is so vigorously challenged by thesauri as an indexing language for special collections.

In addition, investigators at the University of Alberta have felt that, because of the almost universal use and widespread influence of the LC in the Canadian environment and the lack, in Canada, of compensating large scale thesauri or UDC based systems, a thorough investigation of the LC is necessary. Such an investigation may indicate its lack of appropriateness in certain situations or suggest how it may be adapted to perform as an efficient indexing and information language of the type discussed. The detailed basic philosophy underlying the LC investigations at the University of Alberta will not be discussed at length here (19). Only certain preliminary techniques that have been developed and that have been used in this investigation will be briefly outlined.

Work with the MARC tapes and with the LC Subject Heading tapes at the University of Alberta up to this date has produced reformatted MARC tapes used for SDI searching, lists of subject heading words individually related to classifications, and lists of classifications individually related to subject heading words, from both MARC tapes and LC Subject Heading tapes.

For this investigation the list of subject headings from each MARC bibliographic record has been used with the associated LC number. The subject headings from the MARC record have been chosen for the initial tests of the matching procedure described here because it is felt that these represented the true user vocabulary for the set of records and, therefore, could most fairly be regarded as the vocabulary of the indexing language that should be matched with the thesaurus.

We are aware that this approach may be challenged and it is proposed to carry this study further by comparisons with subject heading lists and/or words stripped from the schedules or the indexes as these become available. A related study at the University of Alberta concerned with compression coding of MARC records, when completed, may yield further useful information (20). It is also proposed to examine the complete and linked subject headings of each MARC record for further possible relationships between these words and the LC class number and various thesauri.

Details of Methodology. The approach used in investigating the LC for this project is to take all the main terms and use references of the Water Resources Thesaurus, that is, the same list used in the UDC investigation, and to compare the terms with the subject headings stripped from the MARC tapes. All subject headings identical to the thesaurus entry are stripped off with their associated LC number.

This is done by first running against the MARC tape the programs that have been developed previously at the University of Alberta to strip the LC subject headings with their associated LC number. The resulting file is then sorted according to subject heading entry.

The Water Resources file, which has been sorted for the UDC investigation, is then compared with the sorted LC MARC Subject heading file. This is done by automatically stepping through the thesaurus file and the subject heading file, while comparing them with one another. When a match is found the thesaurus term is written out with the matching subject headings listed below it with the associated LC numbers (Table 3).

The comparison technique employed takes only the first x-1 characters of the term, x being the number of characters in the term; then these are compared with the first x-1 characters of the subject headings. This technique accounts for any plurals that occur in either file, but it also leads to some strange matches, such as CHERT from the thesaurus being matched to CHEROKEE from MARC. Further programming would eliminate the majority of these false hits by checking the last character of the thesaurus term being matched and by then taking precautionary steps if the character is an 'a', an 'i', or a 's'. This was shown to be unnecessary since the false hits were few and obvious.

The file resulting from the comparison is now studied. Only a subset of this file is given in Table 3. There seem to be no definite 'entry points' from the thesaurus to the subject classes. For example, from Table 3, it is apparent that the term "outdoor recreation" leads to several main classes in the LC schedules, GV54, GV182, and TC424. This means that a descriptor from the thesaurus could lead to several subject classes all with very distinct LC numbers with no interconnection between them. No such obvious matches appear as have with the UDC comparisons.

At present, however, these results are regarded as tentative. As has been stated, it is planned to carry the LC investigation much further. More sophisticated matching procedures need to be devised and use should be made of the LC Subject Heading tapes. Nevertheless, if we postulate that MARC subject headings constitute the indexing language vocabulary the lack of pattern in the matches is significant.

TABLE 3

W OIL FIELDS

OIL FIELD FLOODING ADDRESSES, ESSAYS, LECTURES	TN870 .p449 1967AA
OIL FIELD FLOODING WATER-SUPPLY	TD224.T4 A333 NO. 44
OIL FIELD FLOODING	HD9564 .C23 NO. 67
OIL FIELDS ALBERTA	HD9564 .C23 NO. 67
OIL FIELDS PRODUCTION METHODS ADDRESSES, LEC..	TN870 .P449 1967AA
OIL FIELDS TEXAS	TD224.T4 A333 NO.44
OIL FIELDS U.S.	TN872.A5 W38

W OUTDOOR RECREATION

OUTDOOR RECREATION	TC424.C2 A62 1968
OUTDOOR RECREATION CALIFORNIA OROVILLE REGION	GV54.C2 S635
OUTDOOR RECREATION ECONOMICS ASPECTS CALIFORNIA	J87 .C2 1965-1967
OUTDOOR RECREATION ECONOMICS ASPECTS U.S.	SB481 .B4
OUTDOOR RECREATION FLORIDA EVERGLADES	S972.F58 C4
OUTDOOR RECREATION GT. BRITAIN	HD951 .A1W9
OUTDOOR RECBEATION GT. BRITAIN	GV75 .B85 1967
OUTDOOR RECREATION LAW & LEGIS. Virginia	
OUTDOOR RECREATION LAW & LEGIS. U.S.	KF26.I542 1969
OUTDOOR RECREATION MICHIGAN	QE125 .A37 NO. 3
OUTDOOR RECREATION MISSOURI VALLEY	GV54.M8 U5
OUTDOOR RECREATION NEVADA FINANCE	JK8501 .N45 NO.3
OUTDOOR RECREATION NEW HAMPSHIRE	GV54.N4 A2 NO.13
OUTDOOR RECREATION NEW MEXICO	GV54.N6 A5 1968
OUTDOOR RECREATION OREGON	GV54.07 A35 1967
OUTDOOR RECREATION TENNESSEE FINANCE	JK5274 .A27 1968,B13
OUTDOOR RECREATION U.S.	GV53 .A46
OUTDOOR RECREATION VERMONT	GV54.V5 A43
OUTDOOR RECREATION WASHINGTON (STATE)	GV54.W2 A4
OUTDOOR RECREATION WASHINGTON (STATE)	GV54.W2 A5 1968
OUTDOOR RECREATION WISCONSIN	S932.W6 A45
OUTDOOR RECREATION	GV182.2 .D3
OUTDOOR RECREATION	GV182.2 .P7

Tentative Conclusions

The following tentative conclusions have arisen from the work to date:

1. The vocabulary used in the parts of the UDC schedules relevant to the classification of water resources material appears to be up to date. This is indicated by the high percentage of words from the Water Resources Thesaurus that occur in the UDC schedules.

2. An overall set of UDC schedules would be needed to classify water resource material. This is indicated by the proportionality of the matches of the subject headings with terms from the thesaurus.

3. The UDC, combined with a thesaurus, is an effective indexing language for water resources material.

4. In most instances, it appears that plurals are used in the UDC schedules in the same manner as they are used in the thesaurus; whereas there seems to be no such consistency in LC.

5. A concordance, such as that resulting from the thesaurus and the UDC schedules would serve to indicate when sections of the schedules need updating.

6. It appears pending the completion of further study, the UDC could be regarded as a metalanguage which could be used effectively for automated retrieval purposes for water resources management. Results from the AIP/UDC project also support this conclusion.

7. Once the concordance resulting from this study is completed a user will be able to perform more effective searches on mechanized UDC data bases.

8. The concordance will also serve effectively as the heart of an on-line information storage and retrieval system and will be used in both searching and indexing, as far as can be seen at present.

SUMMARY COMMENTS

It is often acknowledged that one of the objections to a simple thesaurus-based indexing language with a consecutive accession number used for document location is that the keyword approach gives inadequate indication of main classes. The keyword relationships consist of very small clusters. In handling information it is sometimes convenient to use both the specialized approach of fine indexing and the broader approach of a classification scheme. The English Electric Thesaurofacet allows this type of combined approach, through an amalgam of faceted indexing and a thesaurus type subject index. This paper discusses a similar linking of thesaurus and schedules.

The technique described in this paper will allow a judgement to be made about the interrelationship of any thesaurus to an existing classification scheme when computer bases are available. If the relationship is satisfactory it will allow the combination of the thesaurus and the classification notation. It offers many aids to the subject specialist who will have to make the decisions about the relationships of the subject descriptions.

In summary, it may be said that developments in computing science, especially those over the past ten years, have given us not only a new way of looking at indexing languages, but also have provided us with new techniques whereby the efficiency or appropriateness of these languages may be less subjectively judged. These judgements of efficiency and appropriateness are only one area now open to those interested in classification and indexing theory. We can also study the change of the indexing language over time, as well as the changes in group concepts about the organization of knowledge. We can build indexing languages, through computer manipulation, which will respond to the changing needs of subject areas. These new insights and techniques, however, will not appear quickly. They require much investigation by people familiar with new theories and new tools.

It should also be added that to gain the most benefit for classification research the applications of theories developed for programming languages and computer systems must also be combined with the results of developments in other areas, such as those in transformational and generative grammars, in the theory of classification (Classification Research Group, Perrault), in the building of mechanized data bases (Rigby, Freeman, Atherton). It seems possible that, at this point in time, enough familiarity and expertise have been developed that theories, some of which were expressed ten years or more ago (21, 22), may now fruitfully be combined to develop useful practice. It is hoped that the project, whose methodology and tentative conclusions are described in this paper, may provide, as it progresses, both practical aids and further theory.

ACKNOWLEDGEMENTS

This research has been partially supported by grants and contracts from the National Research Council of Canada and the Canada Department of Fisheries and Forestry. In addition, the authors gratefully acknowledge the assistance and enthusiastic support of the following: Robert Freeman, Malcolm Rigby, John Batteke, Professor J. Bond, Judith A. Werdel, Hans Wellisch, G.A. Lloyd, F. Alber.

BIBLIOGRAPHY

1. Borko, H.,"Conceptual Foundations of Information Systems," in Frontiers of Librarianship, Syracuse University, 1968, p. 67

2. Freeman, Robert, "Actual and Potential Role of the Universal Decimal Classification in Language Sciences Documentation," in Information in the Language Sciences, New York, Elsevier, 1968, (Proceedings of the Airlie House Conference) p. 149

3. Gardin, Jean-Claude, "On Some Reciprocal Requirements of Linguisitics and Information Techniques," in Information in the Language Sciences, p. 95

4. Batty, C.D.,"The Automatic Generation of Index Languages," Journal of Documentation, 25, 1969, p. 142

5. United States, Office of Naval Research, DOD Manual for Building a Technical Thesaurus,April 1966, AD 633 279

6. United States, National Library of Medicine, Bibliographical Services Division, Medlars Indexing Manual, 1969

7. United States, Library of Congress, Processing Department, Subject Headings Used in the Dictionary Catalogs of the Library of Congress, Washington, 1966, 7th ed.

8. Cleverdon, C.W., "The Cranfield Tests on Index Language Devices," ASLIB Proceedings, 19, 1967, p. 173

9. Lancaster, F.W. "Evaluating the Performance of a Large Operating Retrieval System," in Electronic Handling of Information; Testing and Evaluation, Washington, Thompson, 1967, p. 199

10. Coates, E.J., "Switching Languages for Indexing," Journal of Documentation, 26, 1970, p. 102

11. Stokolov, N.A. and Vleduts, G.E., "Principal Outlines of Information Languages in a Search of Titles for Chemical Publication," Naucho-Techniskaya Informatsiya, 1966, p.19

12. Gottlieb, C.C., "Semantic Clustering of Index Terms," Journal of the Association of Computing Machinery, 15, 1968, p. 493

13. Akiyama, S., Automatic Document Classification Systems, Master's Thesis, University of Alberta, in process (Compares Keyworded Data Base with Manual Classification)

14. Aitchison, Jean, "The Thesaurofacet: A Multipurpose Retrieval Tool," Journal of Documentation, 26, 1970, p. 187

15. Freeman, Robert and Atherton, Pauline, Final Report of the Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Retrieval System, Report AIP/UDC-9, May 1, 1968, American Institute of Physics, New York.

16. Roberts, Brian, Universal Decimal Classification for Use in Polar Libraries, FID no. 348, 1963, 2nd ed.

17. Imroth, J.P., "A Guide to the Library of Congress Classification, Rochester, New York, Libraries Unlimited, 1968

18. The Use of the Library of Congress Classification; Proceedings of the Institute on the Use of the Library of Congress Classification, R.H.Schimmelpfeng and C.D.Cook, ed. Chicago, American Library Association, 1968

19. Heaps, Doreen et al. "Search programs for MARC tapes at the University of Alberta," Western Canada Chapter, American Society for Information Science, Proceedings of the Annual Meeting, Sept. 1970 (Issued by Information Systems, University of Calgary, Dec 1970),p. 83

20. Heaps, H.S. and Reid, W.D., Compression of Data for Library Automation, paper presented to CACUL Conference on Automation in Libraries, June 1971, Vancouver (Sponsored by CACUL Committee on Library Automation)

21. Farradane, J., "Introduction," Information Storage and Retrieval, 1, 1963, p. 1

22. Langleben, M.M. and A.L. Shumilina, "Translating the Titles of Chemical Abstracts into Information Language," Information Storage and Retrieval, 1, 1963, p. 41

INDEXING LANGUAGES
ON ECONOMICS.UDC, UNESCO AND OECD

Maria M.CRUZEIRO
National Institute of Industrial Research Documentation Department Lisbon,Portugal

1. The main purpose of any classification system, either numerical, mixed or non codified, symbolic or not, consists always in grouping concepts in a special field of knowledge, with all its problems.

   Although we have to face these problems, when we are dealing with general subjects, they become more and more complex when we go into the specific fields of a science, which nowadays are often almost a science in themselves.

2. So, we may expect two things from a classification system.

   First, we must find in the chosen system every detail which enables us to distinguish within the main divisions other specified classes. And, if it happens, we may be sure that the given information would be the right one. Such an attitude is the only one which would permit the optimization of the scientific information.

   Second, the classification system must give to the researcher , along with the data refering to the main problem he is studying, also some information concerning marginal problems in connexion with the principal subject.

3. This is the optimal system in economic research because, whenever every researcher begins to study a subject, even if it is familiar to him, he has to think about other problems refering to the main one, before getting the solution. And those subjects may be found in a classification system, within the same divi -

sion - the one which identifies the main subject-or on the contrary - within other divisions, which, by the way, may at first sight not be related with the principal one.

4. Let's see an example.

A group of industrials decide to invest in an extractive industry which involves a special location.

It would also be necessary to solve problems such as - industrial implantation and residential quarters for workers; manpower and wages; health and assistance; pollution and environment, transports, quality control, marketing, productivity, training, etc.

So, if we want to provide the staff in charge of the industrial implementation study with the right documentation they need, we must have, in our Centers, every document classified within a general system like UDC or within several specific systems which can answer every question we are put.

5. When a documentalist has to selected a classification system , he will certainly choose it considering the purpose of the documentation center and, above all, the branch of science it will cover.

But along his work, the documentalist will often use different classifications from that one he had chosen first. On the other hand, he will have to ask information to other Documentation Centers. And, then he will begin to face the problem of the necessity of a universal language to allow all these contacts.

6. Since its creation in 1895 and its dissemination in 1927, the UDC became gradually the connexion among those who work in different documentation centers. Thanks to UDC, we can buy, sell and change analytical cards, systematic catalogues, periodicals and books catalogued according to international standards.

In spite of that, UDC has always had opponents, specially among the users of Documentation Centers, because they are asked to explain minutely what they need, and sometimes they are not able to do it perfectly.

In what concerns documentalists it has been sometimes accused of being vague in some subjects and specially of not being brought up to date with the new events in the technical, scientific, economic and social fields; however UDC system gives to us a quiet great number of details within main divisions.

7. But there are few Documentation Centers which have not a file organized according to the UDC and which we use as an indispensable tool.

A documentalist needs UDC, because there he can find every subject classified within the right division it belongs to. We may also find answers to geographical questions, form of documents, language, race and nationality, and so on. And, if all this is not enough we may use the auxiliary of point of view to give the right classification.

8. But the users can't understand all this (I mean they can't accept it and be preocupated with all this). Everything they want is a simple and precise language to express quickly and distinctly the concept within which is the problem they are worried about.

According to this evolution, two kinds of classification turned up to please both documentalists and users - a specific classification of science and technique separately, which allows a deeper analysis of the documents than when they were analysed together in a general classification; and the classification which gives rise to the "thesauri" and it is, of course, a national one. But this is not very functional, if we think of the universality of progress and of the need of international cooperation.

9. Among the classifications on economics we may speak about the well known classification prepared in 1951 by UNESCO, and which is used in its " International Bibliography of Social Sciences" But a classification must be always brought up to date, for we can't forget that the growth of scientific research is not the same than that of documentation progress.

10. The reason for all this, is that we are so busy with the work in the Centers, preparing documentation and giving information to the researchers that we haven't any time to look at our own problems.

    But we can say that a new system appeared, which can follow the progress of scientific research. It can be compared to the human life, because like the dead cellules are rejected and the new ones take their place, so the useless concepts are replaced by the new ones.

11.. This system is not anything new, it is the use of the basic and structural principles of a universal classification, adapted in a free way.

    As an example of this system, the Lists of Descriptors emerged and among them the "Alligned List of Descriptors on Social and Economic Development" prepared by ILO (International Labour Organization) DSE (Deutsche Stiftung für Entwicklungsländer) FAO (Food and Agricultural Organization) and OECD (Organization for Economic Cooperation and Development).

    The above mentioned list has been used by the Documentation Centers in those countries interested in international exchange of documents and information.

12. The main purpose of this of descriptors is to permit an analysis of the documents as complete as possible and at the same time to allow the conversion of this indexing language into a language to computer.

This list of descriptors is kept up to date, thanks to the cooperation among those centers, which have collaborated withe OECD submitting to its judgement the new descriptors they find necessary.

13. Such a system is unique, for it is constantly brought up to date.

The worst disadvantage of any classification system is the possibility of becoming useless, but with the descriptors we may say that this problem is solved.

However, this solution has to be faced by both documentalists and users, because, although they use the same language, they use it with different purposes, that mans, with complemantary objectives.

14. There is a field that concerns only the documentalists, it is the storage of information. If this storage is prepared in computer, so the documentalist must give a look at all documents classified within different systems of classification and convert them into a language, which can be used in a mechanized system.

Documentalists may use the basic structure of UDC as a reference, to prepare this conversion.

So, the indexing language by key-words represents an intermediate step between a general classification (like UDC) and its mechanization.

15. In Portugal, we begin to realize the necessity of sources of economic information, in the last years. And so, the most important private enterprises and several departments of the government which are preparing the development planning and studies on macroeconomics decided to create Documentation Centers to provide their technical staff with the information they need.

Since the begining of last year, we started a movement of cooperation in the field of Economic and Social Documentation. The first work prepared by this group is the outcoming "Catalogue of Periodicals on Economic and Social Sciences", covering all the periodicals belonging to the components of this working group. The catalogue contains references of 3.300 periodicals and is divided into three parts:

I - Alphabetical arrangement by titles.

II - Alphabetical arrangement by subjects.

III - Alphabetical arrangement by location and international organizations.

We have in mind a permanent actualization of this catalogue with a new number every year and a supplement each six months.

This is the first step of a program of work to be undertaken by the members of this "pool" . But we want to go further.We intend to explore every publication cover to cover, according to the needs of the department interested in the collaboration, to prepare the analysis of the caticles in the Portuguese periodicals, and also to buy international abstract publications covering foreign texts on economics and social sciences.

We intend also to publish systematically all these references in a Bulletin of Documentation, in specialized bibliographies, in permanent information, and whenever we will be asked.

16. So, in Portugal, the documentalists working in the economic and social field have a great interest in knowing about any possibility of relation between interest UDC and other classification systems and their possible mechanization.

А.И.Черный (СССР)

О ЛОГИЧЕСКИХ ПРИНЦИПАХ ПОСТРОЕНИЯ УДК И ПОИСКОВЫХ ЯЗЫКОВ ДЕСКРИПТОРНОГО ТИПА

Цель данной работы - показать, что хотя дескрипторный поисковый язык (ДПЯ), как и УДК, представляет собой классификацию, предназначенную для поиска текстов (документов) по их предметному и/или тематическому содержанию, эта классификация основана на совершенно иных логических принципах, чем УДК. Данные различия обусловлены тем, что УДК является в значительной мере классификацией предкоординированного, перечислительного типа (здесь мы имеем в виду основные классы УДК), тогда как ДПЯ - классификация посткоординируемого типа.

Если будет доказано, что ДПЯ (далее мы будем называть их дескрипторными классификациями) и УДК построены действительно на разных логических принципах, то это позволит нам сделать ряд важных выводов принципиального характера, которые можно будет распространить и на другие классификации перечислительного типа, а именно - на библиотечно-библиографическую классификацию Государственной библиотеки СССР им.В.И.Ленина, на Десятичную классификацию Дьюи, на Классификацию Библиотеки конгресса США, а также на алфавитно-предметные классификации. Эти выводы заключаются в следующем.

I. Каждый тип классификации, предназначенный для поиска документов и/или предметов, имеет такую логическую структуру, которая обусловлена

- типом поисковых задач, решаемых с помощью этой классификации;

- способом ее использования;
- особенностями технических средств, которые применяются для осуществления поиска.

Поэтому использование классов УДК для координатного индексирования документов - а для такого способа использования классы УДК при ее построении не предназначались - должно приводить к значительно худшим результатам, чем использование дескрипторной классификации.

2. Встречается еще немало специалистов, работающих в области информационного поиска, которые считают,что все различия между иерархическими, фасетными, алфавитно-предметными и дескрипторными классификациями заключаются в способе их использования для индексирования документов, и что между этими типами классификаций нет существенных логико-лексических различий. Такая точка зрения основана на глубоком заблуждении, что алфавитно-предметная, иерархическая, фасетная и дескрипторная классификации по одному и тому же предмету состоят из одних и тех же классов (имен), но только эти классы упорядочены в каждой из данных классификаций специфическим образом.

В этом заблуждении таится причина того, что появляются труды о перспективности использования УДК в режиме координатного индексирования /I/, о "зеркальности" иерархических и дескрипторных классификаций /2/ и т.д. И хуже всего то, что для обоснования таких мнений иногда проводятся обширные эксперименты, которые могут продолжаться по два-три года и требуют значительных расходов. Пухлые отчеты о таких экспериментах, а также журнальные статьи об их результатах обычно содержат в изобилии таблицы, графики и математические формулы, которые у неискушенного читателя могут создать впечатление о высокой "научности" этих трудов. Нам хотелось бы посоветовать таким читателям никогда не забывать той простой истины, что если исходная гипотеза, для подтверждения которой проводится эксперимент, недостаточно обоснована или даже неверна, а методика проведения такого эксперимента плохо разработана, никакие математические ухищрения не могут подтвердить то, чего нет в действительности.

"Исследования" такого типа наносят большой вред разработке теории поисковых систем , т.к. создают путаницу в понятиях и терминологии, а также порождают опасную иллюзию того, что в теории поисковых систем все уже более или менее ясно, что в ней нет сколько-нибудь сложных логико-семантических проблем, которые нам еще предстоит решить; наконец, что дальнейшие исследования в этой области являются в значительной степени беспредметными. Тем самым задачи, решаемые при разработке теории поисковых систем, низводятся до уровня тривиальных, а специалисты отвлекаются как от глубокого исследования логико-семантических принципов построения и использования дескрипторных классификаций - с целью повышения эффективности последних, так и от дальнейшего совершенствования методики использования УДК, в чем ныне ощущается все более острая практическая потребность. Вместо этого специалистов толкают на путь необоснованного и бесплодного экспериментирования, которое иногда весьма смахивает на занятия достославных академиков из Лагадо.

3. УДК – классификация полуфасетного типа. Это уже было показано нами в одной из предшествующих работ /3/. Совершенствование УДК должно осуществляться главным образом за счет более полного использования комбинационных возможностей, содержащихся в этой классификации, а также за счет разработки более совершенной методики ее использования и организации быстрого переиздания таблиц УДК (например, путем применения либо ЭЦВМ в сочетании с быстродействующими фотонаборными машинами, либо устройств типа "Композ-лайн", "Фотолист" и "Листоматик" для автоматического фотографирования машинописных строк текста, нанесенных на карточки формата I87x82,5 мм) Необходимо соблюдать предельную осторожность в любых нововведениях в УДК, т.к. они могут оказаться разрушительными для этой классификации.

4. Вопрос об использовании вычислительных машин для осуществления информационного поиска на основе УДК не имеет никакого отношения к обсуждаемой здесь теме. Это чисто экономический вопрос, который должен решаться применительно к конкретным условиям.

Для того чтобы доказать правильность сформулированных нами выводов, рассмотрим основные логические принципы построения дескрипторных классификаций. С этой целью определи предварительно следующие термины.

Классификация – система классов, предназначенная для поиска документов или предметов по тем или иным признакам.

Классифицирование – процесс распределения документов или предметов по классам на основании признаков, присущих этим документам или предметам.

Документ - любой записанный на каком-то материальном носителе осмысленный текст, который обладает определенной логической завершенностью и содержит сведения об его источнике и/или создателе. По этому определению документом является не только книга, статья и т.д., но и ее отдельный фрагмент - глава, раздел, абзац.

Индексирование - процесс классифицирования, дополненный операцией обозначения документа или предмета именем (индексом) класса, в который включен данный документ или предмет.

Координатное индексирование - способ индексирования, который характеризуется тем,что основное смысловое содержание какого-либо документа или отличительные признаки какого-либо предмета выражаются в виде логического произведения соответствующих классов, обозначенных словами естественного языка или их кодами, причем это делается с целью последующего отыскания данного документа или предмета по тем или иным признакам среди множества других.

Основное смысловое содержание - центральная тема или предмет, обсуждаемый автором в документе. Обычно основное смысловое содержание документа более или менее точно и полно выражается в заглавии этого документа.

Дескрипторная классификация - система классов, предназначенная для координатного индексирования документов и/или предметов. Этот термин мы считаем синонимом термина "дескрипторный поисковый язык".

Дескриптор - класс или имя класса дескрипторной классификации. В данной работе термины "дескриптор", "унитерм" и "ключевое слово" мы будем считать синонимами, хотя они выражают понятия, отличающиеся по объему друг от друга.

Поисковый образ - присвоенная какому-либо документу или предмету совокупность имен классов или их шифров, которая выражает основное смысловое содержание этого документа или отличительные признаки этого предмета и предназначена для отыскания данного документа или предмета среди множества им подобных.

Поисковое предписание - совокупность имен классов или их шифров, которая выражает важнейшие отличительные признаки искомых документов или предметов и служит для их отыскания среди множества других - путем сравнения с поисковыми образами документов или предметов данного множества. В поисковом предписании могут также быть указаны дополнительные условия, лишь при соблюдении которых поисковый образ может считаться эквивалентным поисковому предписанию.

Посткоординация - способ построения сложных классов, в которые мы при классифицировании включаем документы или предметы, характеризующийся тем,что такие классы строятся непосредственно в момент классифицирования (индексирования) каждого данного документа или предмета, а не предварительно, до его классифицирования, т.е. не при построении самой классификации, как это делается в предкоординированных, перечислительных классификациях. Построение таких сложных классов производится путем логического перемножения простых, элементарных классов, соответствующих отличительным признакам классифицируемых документов или предметов.

Семантическая сила - относительная способность классификации обеспечивать исчерпывающее, точное и однозначное индексирование документов с любым смысловым содержанием или любых предметов; определяется относительно естественного языка, который обладает максимальной семантической силой.

Необходимо отметить, что в поисковых образах документов или предметов и в поисковых предписаниях дескрипторы никак не связаны между собой (точнее - связаны лишь отношением координации) и могут функционировать независимо друг от друга. Логическое перемножение понятий, выражаемых ими, производится дважды: при координатном индексировании документа или предмета и при сравнении поискового образа с поисковым предписанием.

Если бы это утверждение было неправильным, то была бы невозможна инвертированная организация поисковых массивов, при которой поисковые образы расчленяются на составляющие их дескрипторы. Кроме того, было бы невозможным использование - в качестве критерия соответствия документа или предмета запросу-не только полного совпадения поискового образа с поисковым предписанием, но и полного вхождения поискового предписания в поисковый образ. Это означает, что при использовании дескрипторных классификаций поисковые образы не обладают логической целостностью, т.к. в противном случае самостоятельное использование их отдельных частей было бы недопустимо.

Данное утверждение справедливо и тогда, когда в дескрипторных классификациях используются элементы грамматики - указатели связи и/или указатели роли. Обозначенные ими синтагматические отношения между дескрипторами существуют в поисковых образах потенциально и актуализируются только в момент сравнения поискового образа с поисковым предписанием.

Дескрипторные классификации предназначены для поиска документов или предметов по любому, заранее не предвиденному сочетанию признаков. Это достигается путем применения метода координатного индексирования. Важнейшие особенности данного метода заключаются в следующем.

I. Для координатного индексирования документов или предметов используются классы, не связанные друг с другом никаким другим отношением, кроме отношения принадлежности к одному общему классу (множеству).

2. При координатном индексировании производится логическое перемножение классов, которые обозначены (названы) словами естественного языка.

3. Логическое перемножение классов производится многократно: сначала - при координатном индексировании документа или предмета, а затем-при каждом сравнении поискового образа этого документа или предмета с поисковым предписанием.

4. В поисковых образах и поисковых предписаниях имена классов связаны друг с другом лишь отношением вхождения в одно подмножество.

Из этих особенностей координатного индексирования вытекают основные требования к логическим принципам построения классов, которые должны составлять дескрипторную классификацию. Задача заключается в том, чтобы отобранные для нее классы обеспечивали достижение следующих целей:

а) Возможность построения любых отсутствующих классов (понятий) способом логического перемножения имеющихся.

б) Однозначность толкования порожденных таким способом классов.

в) Минимальное число классов в дескрипторной классификации.

Как было ранее отмечено, дескрипторная классификация не имеет ни одного уровня иерархии. Это означает, что в ней мы оперируем с простыми классами, которые не содержат видовых подклассов (или – по крайней мере – выступают как классы, не имеющие таких подклассов). Такие простые классы мы будем называть далее элементарными.

Что же такое постулируемый нами элементарный класс?

Определение "элементарный" заставляет нас предположить, что такими классами должны быть какие-то минимальные классы, дальнейшее логическое деление или расчленение которых невозможно без потери определенного качества. Условимся, что далее мы будем употреблять термины "понятие" и "класс" как синонимы, хотя термин "класс" выражает только одну сторону понятия – его объем. Кроме того, будем четко различать простое расчленение понятий от их логического деления. При правильном логическом делении подстановка любого из членов деления вместо x в формулу $P(x)$ всегда дает истинные суждения, а подстановка любого из членов расчленения – ложные).

В дескрипторной классификации элементарным не может быть предельно узкий класс, соответствующий единичному понятию. Если бы это было так, то для построения классов, не содержащихся в данной классификации, нам пришлось бы представлять каждый такой класс в виде логического произведения, состоящего из многих единичных понятий. А это сильно увеличивало бы вероятность образования ими паразитных и неясных понятий. Мы называем паразитными такие понятия, которые не совпадают с образуемым, но соответствуют реально существующим понятиям. Кроме того, тогда в дескрипторной классификации будет содержаться слишком много классов, что существенно затрудняет ее практическое использование.

Например, возьмем единичные понятия ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ ДЬЮИ, УНИВЕРСАЛЬНАЯ ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ и БИБЛИОТЕЧНО-БИБЛИОГРАФИЧЕСКАЯ КЛАССИФИКАЦИЯ ГОСУДАРСТВЕННОЙ БИБЛИОТЕКИ СССР им.В.И.ЛЕНИНА и КЛАССИФИКАЦИЯ БИБЛИОТЕКИ КОНГРЕССА США. Если логически перемножать эти понятия,то образуется более широкое понятие,которое можно толковать и как ИЕРАРХИЧЕСКАЯ КЛАССИФИКАЦИЯ, и как БИБЛИОТЕЧНО-БИБЛИОГРАФИЧЕСКАЯ КЛАССИФИКАЦИЯ. Такая неопределенность возникает по той причине, что в выражающих эти понятия терминах эксплицитно не выражен какой-то общий для всех них родовой признак. По-видимому, логическое перемножение единичных понятий вообще не допустимо.

С другой стороны, в качестве элементарных классов далеко не всегда можно использовать и предельно широкие классы категориального характера. При построении понятий, соответствующих таким широким классам, в содержании этих понятий не сохраняются видовые признаки входящих в них понятий. Если процесс обобщения понятий довести до достаточно высокого уровня, то мы можем потерять возможность строить из них - путем логического перемножения - узкие однозначные понятия. Это можно показать на следующем примере.

Допустим,что у нас имеется понятие САМОЛЕТЫ. Это понятие можно заменить более широким понятием ЛЕТАТЕЛЬНЫЕ АППАРАТЫ. При такой замене из содержания понятия САМОЛЕТЫ были исключены его отличительные признаки "быть тяжелее воздуха", "иметь двигатели" и "иметь несущие крылья". Если в дескрипторной классификации не будет содержаться классов, соответствующих этим признакам, то тогда мы не сможем построить класс "пассажирские самолеты", пользуясь лишь классами ЛЕТАТЕЛЬНЫЕ АППАРАТЫ И ПАССАЖИРЫ.

Теперь попытаемся дать определение элементарного класса. A priori очевидно, что любое определение такого чрезвычайно тонкого, хотя интуитивно - казалось бы - в достаточной мере ясного понятия, будет весьма условным, содержащим немало исключений. Но мы убеждены, что наличие даже такого ориентирующего определения может принести большую пользу при создании дескрипторных классификаций, т.к. создатели таких классификаций вначале полагаются в основном на свою интуицию, а затем долго корректируют первоначально построенную дескрипторную классификацию на основании результатов ее практического использования.

Пусть дано понятие САМОЛЕТЫ, которое можно разделить на более простые понятия, например, следующим образом.

САМОЛЕТЫ=ЛЕТАТЕЛЬНЫЕ АППАРАТЫ & НЕСУЩИЕ КРЫЛЬЯ
& ДВИГАТЕЛИ

(Здесь и далее символ " & " обозначает операцию логического умножения понятий).

Необходимо отметить, что такая схема разделения понятия САМОЛЕТЫ не является произвольной. При построении иерархических классификаций каждое понятие (класс) разделяется на видовые понятия (подклассы). Здесь использован совсем другой принцип, который ближе к операции расчленения понятия. При разделении понятия мы заменяем его определением, состоящим из родового понятия и одного или нескольких более узких понятий, призванных выполнять функцию видового отличия. Например:

I) ЦВЕТНЫЕ КИНОФИЛЬМЫ = КИНОФИЛЬМЫ (K_I)
& ЦВЕТ (K_2)

2) РЕАКТИВНЫЕ ПАССАЖИРСКИЕ САМОЛЕТЫ = САМОЛЕТЫ (K_I) &
ПАССАЖИРЫ (K_2) & РЕАКТИВНЫЕ ДВИГАТЕЛИ (K_3)

В этих примерах родовым понятием (классом) является K_I, а функцию видового отличия выполняют понятия K_2 в первом примере и K_2 и K_3 - во втором.

Попробуем реконструировать исходное понятие САМОЛЕТЫ путем логического перемножения понятий ЛЕТАТЕЛЬНЫЕ АППАРАТЫ, НЕСУЩИЕ КРЫЛЬЯ И ДВИГАТЕЛИ. Для этого будем поочередно рассматривать каждое из этих понятий в качестве опорного (родового), а остальные - в качестве пояснительных, видовых. В результате выполнения такой процедуры мы увидим, что из понятий ЛЕТАТЕЛЬНЫЕ АППАРАТЫ, НЕСУЩИЕ КРЫЛЬЯ И ДВИГАТЕЛИ получается не только исходное понятие САМОЛЕТЫ, но и паразитное понятие АВИАЦИОННЫЕ ДВИГАТЕЛИ. Это позволяет сделать вывод, что разделение понятия САМОЛЕТЫ на образующие его понятия (по схеме определения этого понятия через указание рода и видового отличия) не допустимо. Следовательно, понятие САМОЛЕТЫ можно считать элементарным.

Теперь дадим операциональное определение элементарного класса. Если понятие, соответствующее какому-либо классу, логически разделить (расчленить) на составляющие его понятия по схеме определения этого понятия через указание рода и видового отличия, а затем вновь логически перемножить все эти понятия, принимая поочередно каждое из них в качестве опорного (родового), то мы получим либо одно, и только одно, исходное понятие, либо дополнительно одно или несколько других паразитных понятий. Когда мы получаем одно, и только одно, исходное понятие, то оно не является элементарным и – следовательно – может быть разделено на другие понятия. Когда же логическое перемножение составляющих понятий дает два или более неэквивалентных понятий, то тогда исходное понятие можно считать элементарным. По-видимому, при разделении элементарного понятия в его содержании теряется что-то такое, что уже не позволяет нашему сознанию однозначно реконструировать его из составных частей. Такое операциональное определение можно задать и в виде последовательности действий, которая показана на рис.I.

Применение только что описанной процедуры для проверки элементарности понятий (классов) не всегда необходимо. Например, элементарность абстрактных понятий типа АНАЛИЗ, ВРЕМЯ, ИСПОЛЬЗОВАНИЕ, КОЛИЧЕСТВО, СКОРОСТЬ, ЦВЕТ, ЭФФЕКТИВНОСТЬ и т.д. очевидна из без такой проверки. Кроме того, не требуют проверки на элементарность и некоторые другие неабстрактные понятия, например, КРЫЛО, КОЛЕСО, ЛУЧ, СУДА и т.д. Но если возникает сомнение в элементарности какого-либо из неабстрактных понятий, то может оказаться достаточной попытка логически разделить (расчленить) это понятие по схеме его определения через указание рода и видового отличия. Если решение такой задачи окажется затруднительным, то этого уже может быть достаточно, чтобы считать проверяемое понятие (класс) элементарным. В продуктивности такого приема, можно убедиться, применив его, например, к понятию КРЫЛО.

Введение понятия элементарного класса и его операциональное определение, как нам кажется, существенно дополняет определение дескриптора как класса условной эквивалентности и как номинального определения понятия, соответствующего такому классу. Номинальным мы называем определение, представляющее собой описательное выражение основного содержания понятия, которое приписывается в качестве смысла всем и каждому имени, включенным в данный класс условной эквивалентности. Понимание дескриптора как класса условной эквивалентности (или как имени такого класса) позволяет нам группировать близкие по смыслу слова и словосочетания естественного языка в классы и таким образом придавать дескрипторам необходимую однозначность. Подход к дескриптору, как к номинальному определению понятия, выражаемого этим дескриптором, помогает нам уточнить его смысл. А определение дескриптора как имени элементарного класса в том смысле, как это было разъяснено ранее, помогает нам правильно выбирать логический уровень слов и словосочетаний, используемых в качестве дескрипторов.

Использование понятия элементарного класса позволяет также несколько прояснить существо трудностей, возникающих при переводе дескрипторных классификаций с одного естественного языка на другой. Как известно, в каждом естественном языке содержатся слова, которые не имеют эквивалентных им по смыслу слов в других языках. Например, в русском языке нет слов, которые взаимно однозначно соответствовали бы английским словам bit, car, vehicle, hardware, software и т.д. Поэтому если в качестве имени элементарного класса (дескриптора) будет выбрано слово или словосочетание, которое не имеет точного смыслового эквивалента в словарном составе какого-то другого языка, то при переводе такого слова или словосочетания на данный язык оно либо заменяется той или иной совокупностью из нескольких слов,либо передается описательным путем. Естественно, что в результате этого разрушаются построенные для исходного языка классы условной эквивалентности, а также изменяется логический уровень имен таких классов. Вместо прежних классов строятся новые элементарные классы, соответствующие особенностям основного словарного состава данного языка. Таким образом, перевод дескрипторной классификации с одного языка на другой (в строгом понимании смысла термина "перевод") оказывается практически невозможным, что подтверждается целым рядом известных примеров.

Сегодня еще очень трудно оценить как преимущества, которые может дать использование идеальных дескрипторных классификаций, так и трудности, которые стоят на пути построения таких идеальных дескрипторных классификаций, содержащих только элементарные классы. Но интуитивно ясно, что чем больше реальная дескрипторная классификация будет похожа на идеальную ,тем больше преимуществ мы получим при ее использовании.

Идеальная дескрипторная классификация должна быть похожа на "минимальный словарь" Б.Рассела, который дал этому термину такое толкование /4/:

"Как правило, существует несколько способов, с помощью которых слова, употребляемые в науке, могут быть определены небольшим числом терминов из числа этих слов.Эти немногие термины могут иметь или наглядные, или номинальные определения с помощью слов, не принадлежащих к данной науке, или - пока наука не "интерпретирована" ... - могут быть оставлены как бы без наглядного и без номинального определения и рассматриваться просто как набор терминов, имеющих те свойства, которые наука приписывает своим основоположным терминам. Такой набор начальных слов я называю "минимальным словарем" данной науки, если только (а) каждое иное слово, употребляемое в науке, имеет номинальное определение с помощью слов этого минимального словаря и (в) ни одно из этих начальных слов не имеет номинального определения с помощью других начальных слов".

Построение дескрипторной классификации из элементарных классов позволяет иметь в ней минимальное число классов (минимальный словарь), что значительно упрощает процедуру координатного индексирования. Кроме того, сокращение числа неэлементарных классов в дескрипторной классификации позволяет расширить ее комбинационные возможности, а - следовательно - увеличить семантическую силу такой классификации. Это можно показать на следующем примере.

Пусть нам даны следующие классы:

K_I - САМОЛЕТЫ
K_2 - КРЫЛЬЯ
K_3 - СТРЕЛОВИДНАЯ ФОРМА
K_4 - ПАССАЖИРЫ
K_5 - СВЕРХЗВУКОВАЯ СКОРОСТЬ

Тогда путем логического перемножения разных сочетаний из этих классов можно получить следующие более сложные классы:

K_1 & K_2 = КРЫЛЬЯ САМОЛЕТОВ
K_1 & K_4 = ПАССАЖИРСКИЕ САМОЛЕТЫ
K_1 & K_5 = СВЕРХЗВУКОВЫЕ САМОЛЕТЫ
K_2 & K_3 = СТРЕЛОВИДНЫЕ КРЫЛЬЯ
K_2 & K_5 = КРЫЛЬЯ ДЛЯ СВЕРХЗВУКОВЫХ СКОРОСТЕЙ
K_1 & K_2 & K_3 = СТРЕЛОВИДНЫЕ КРЫЛЬЯ САМОЛЕТОВ
K_1 & K_2 & K_4 = КРЫЛЬЯ ПАССАЖИРСКИХ САМОЛЕТОВ
K_1 & K_2 & K_5 = КРЫЛЬЯ СВЕРХЗВУКОВЫХ САМОЛЕТОВ
K_1 & K_4 & K_5 = СВЕРХЗВУКОВЫЕ ПАССАЖИРСКИЕ САМОЛЕТЫ
K_2 & K_3 & K_5 = СТРЕЛОВИДНЫЕ КРЫЛЬЯ ДЛЯ СВЕРХЗВУКОВЫХ СКОРОСТЕЙ
K_2 & K_2 & K_3 & K_4 = ПАССАЖИРСКИЕ САМОЛЕТЫ СО СТРЕЛОВИДНЫМИ КРЫЛЬЯМИ
= СТРЕЛОВИДНЫЕ КРЫЛЬЯ ПАССАЖИРСКИХ САМОЛЕТОВ
K_1&K_2&K_3&K_4&K_5 = СВЕРХЗВУКОВЫЕ ПАССАЖИРСКИЕ САМОЛЕТЫ СО СТРЕЛОВИДНЫМИ КРЫЛЬЯМИ

Теперь допустим, что вместо классов K_1 и K_5 мы будем использовать класс $K_6 = K_1$ & K_5 (СВЕРХЗВУКОВЫЕ САМОЛЕТЫ). Нетрудно убедиться, что число возможных классов, которые можно построить путем логического перемножения – в разных сочетаниях – классов K_2, K_3, K_4 и K_6, составит 4 вместо 13 до такой замены. Попробуем перемножить другие классы – K_2 и K_3 – и вместо них использовать класс K_7 (СТРЕЛОВИДНЫЕ КРЫЛЬЯ). Это приведет к уменьшению числа возможных классов до 8. Наконец, можно перемножить также классы K_1 и K_4 и получить класс K_8 (ПАССАЖИРСКИЕ САМОЛЕТЫ).

Тогда общее число возможных классов, которое можно будет построить из классов $К_2$, $К_3$, $К_5$ и $К_8$, составит тоже 8. Из приведенных примеров видно, что замена в дескрипторной классификации элементарных (или близких к элементарным) классов неэлементарными ведет к снижению комбинационных возможностей таких классов при координатном индексировании.

Таким образом, для координатного индексирования документов или предметов следует использовать не любые слова и словосочетания естественного языка, а лишь имена элементарных классов. И хотя эта рекомендация полностью выполнима лишь теоретически, в идеальной дескрипторной классификации, мы не должны забывать о ней при построении реальных дескрипторных классификаций, а также при координатном индексировании документов ключевыми словами, выбираемыми непосредственно из текстов этих документов.

Понятие элементарного класса не было чуждо К.Муэрсу и М.Таубе - основным создателям метода координатного индексирования и дескрипторной классификации, труды которых стали незаслуженно забываться сегодняшними специалистами по поисковым системам. Именно К. Муэрс еще в 1950 г. ввел в научное обращение такие термины, как "информационный поиск" (information retrieval),"информационно-поисковая система" (information retrieval system), "дескриптор" (descriptor), "дескрипторный словарь" (descriptor dictionary),"поисковый образ "(delineating set) , "поисковое предписание" (prescribing set) и ряд других.

В одной из своих ранних работ (1950г.) К. Муэрс писал: "Предмет каждого документа или другого элемента информации характеризуется или описывается с помощью некоторого множества "дескрипторов", взятых из формального словаря таких терминов. В качестве грубого приближения к тому, что здесь имеется в виду, на память приходит "список предметных заголовков "(subject heading list) / 9, с.20/. И далее: "Каждый дескриптор этого множества применим или в каком-то смысле истинен в отношении предметного содержания элемента информации. При таком типе описания (characterization) дескрипторы функционируют независимо друг от друга. Тот факт, что данное множество состоит из нескольких дескрипторов, может означать,что в исходном документе они образуют какое-то взаимодействующее сочетание; это вполне может также означать, что они относятся к разным самостоятельным идеям, рассеянным в [тексте] документа" / 9, с.26/.

Двенадцать лет спустя К.Муэрс говорил:"Выступающие в качестве дескрипторов понятийные единицы (conceptual units) могут выражаться одним или несколькими словами.В основе дескрипторов лежат элементарные (unit) понятия, а не простые слова" / I0, с.25/. "Дескриптор .... состоит из двух частей: "ярлыка" (label) и устного или письменного "определения" / I0, с.26/.

Идею элементарного класса выражает и термин "унитермы", который был введен М.Таубе. "Унитермы - писал известный американский специалист по поисковым системам Дж.Костелло,- это предметные заголовки, выражающие элементарные понятия (unit concepts). Все унитермы имеют одинаковый иерархический ранг: ни один из них в логическом смысле не стоит выше или ниже какого-либо другого (как в [иерархической] классификации) и ни один не используется в заранее установленном сочетании с каким-либо другим (как в традиционных предметных заголовках) /7/.

Сейчас уже ясно, что К. Муэрс недооценивал роль плана выражения в дескрипторных классификациях и акцентировал внимание в основном на плане содержания, на понятийной стороне дескрипторов. Наоборот, М.Таубе больше внимания уделял плану выражения понятий, используемых для координатного индексирования документов, их словесной оболочке. Это нашло свое отражение даже в том, что он назвал имена дескрипторных классов "унитермами", т.е. элементарными терминами.

Подход М.Таубе был более прагматическим, т.к. при построении дескрипторных классификаций и при координатном индексировании мы оперируем со словами и словосочетаниями естественного языка, что создает немало специфических трудностей. По-видимому, именно такой подход и обеспечил значительно большее распространение идеям М.Таубе, которые были восприняты многими слишком поверхностно и упрощенно, хотя М.Таубе постоянно подчеркивал, что в основе его системы классификации лежит ассоциирование не слов, а идей / I2,I3,I4/. Это отрицательно сказалось на разработке логико-семантических основ построения дескрипторных классификаций.

Теперь кратко рассмотрим логические принципы построения УДК, точнее - ее основной части. Эта задача значительно облегчается тем, что в предшествующей части работы мы уже рассмотрели подробно основные логические принципы, построения дескрипторных классификаций.

Создатели УДК ставили перед собой задачу заблаговременно построить для каждого сочетания признаков отдельный класс, в который можно было бы включать все документы, характеризующиеся данным сочетанием признаков. Такую задачу можно было разрешить лишь путем последовательного деления каждого класса по одному основанию на взаимоисключающие подклассы. Кроме того, требовалось создать такую классификацию, которую не нужно было бы радикально перестраивать каждые 10-15 лет. Поэтому при построении УДК естественно было полагаться не на изменчивое содержание документов, подлежащих классифицированию, не на столь же изменчивые информационные запросы, а на сами научные знания, которые выражаются в документах и информационных запросах. К сожалению, в качестве основы для построения УДК была взята прагматическая Десятичная классификация Дьюи, которую многие - как справедливо заметил Е.И.Шамурин - неправильно считают бэконовской /5/. Эта классификация не соответствовала достигнутому человечеством уровню знаний уже тогда, когда ее составил М.Дьюи (1876 г.) Именно эта схема создала из УДК "прокрустово ложе", в которое уже свыше 50 лет мы тщетно пытаемся втиснуть непрерывно развивающиеся и изменяющиеся человеческие знания.

В основе УДК лежит логическое деление понятий (классов) по видоизменению (разновидностям) определенного признака. Объем и имена классов в УДК не выбираются, а целиком определяются содержанием и объемом делимого класса (понятия). Иначе говоря, при построении УДК в ней вообще нет проблемы выбора логического уровня классов, предназначенных для индексирования документов. Между подразделениями УДК существуют иерархические отношения, которые полностью отсутствуют в дескрипторных классификациях.

В обобщенном виде формально-логические правила построения иерархической классификации можно записать следующим образом:

1) $k_I \cup k_2 \cup K_3 \cup \ldots \cup k_n = K$

2) $k_I \cap k_2 = \emptyset$, $k_2 \cap k_3 = \emptyset$, $k_I \cap k_3 = \emptyset$ и т.д.
– для любой пары подклассов.

Здесь К – некоторый класс, который на основе какого-то одного, и только одного, признака разделен на подклассы $K_I, k_2, k_3, \ldots, k_n$.

Построение элементарных классов (в нашем понимании) не было и не могло быть определяющим принципом при построении УДК, т.к. ее классы предназначались для координатного индексирования документов. Поэтому в УДК содержится лишь небольшой процент элементарных классов. Большинство же классов УДК являются неэлементарными. Следовательно, если использовать многочисленные имена классов УДК для координатного индексирования документов, то комбинационные возможности таких классов будут весьма низкими, а неоднозначность логических произведений будет достаточно большой как по содержанию, так и по форме. Неоднозначностью по содержанию мы здесь называем порождение паразитных понятий в результате логического перемножения понятий, соответствующих классам УДК. А под неоднозначностью по форме мы понимаем возможность построения одного и того же сложного понятия (класса) путем логического перемножения разных сочетаний исходных классов. Это означает, что два документа, посвященные одному и тому же предмету или теме, могут иметь разные поисковые образы.

Например, в УДК для обозначения одного и того же понятия АВТОМАТИЧЕСКИЕ СИСТЕМЫ УПРАВЛЕНИЯ (АСУ) в настоящее время используются следующие индексы: 62-52; 65.0I2.45; 658.5I; 68I.3; 658.0I4.I.0II.5I.0I2.I22:68I.3:330.II5:338.94; 62-50:007:65.0II.56; 65.0I2.I22; 68I.3:658.5I.0II.56; 658.I55:658.5.0II.56; 68I.32:65.0II.56; 658.I.0I2.I22; 658.5I.0II.56 и т.д. Где уж тут требовать однозначного индексирования!

Использование классов УДК для координатного индексирования можно сравнить с попыткой точно выражать сложные мысли на английском языке лишь путем механического комбинирования готовых фраз, содержащихся в русско-английском разговорнике. Число готовых фраз в таком "разговорнике", как УДК очень велико, что создает множество дополнительных затруднений семантического и прагматического характера. В полном варианте УДК только один класс 6 " Прикладные науки. Медицина. Техника" содержат 64 тыс.подразделений, тогда как "Тезаурус технических терминов", который был составлен и издан в США (I964г.), насчитывает всего 7750 дескрипторов /6,II/. При этом можно с уверенностью утверждать, что семантическая сила дескрипторной классификации , содержащейся в этом тезаурусе, значительно выше, чем семантическая раздела 6 УДК . Правда, некоторая часть подразделений класса 6 УДК относится к медицине, которая не нашла отражения в "Тезаурусе технических терминов". Но данный тезаурус охватывает еще значительную часть класса 5 УДК - "Математика. Естественные науки" , в котором насчитывается 35 тыс.подразделений.

Итак, возможно ли использование классов УДК для координатного индексирования документов ? На этот вопрос мы даем такой ответ: в принципе - возможно, но на практике этого делать не следует, т.к. классы УДК не предназначены для такого способа использования. С другой стороны, в этом нет никакой необходимости, т.к. эффективность УДК весьма высока, если ее использовать умело, в свойственном этой классификации режиме и для решения поисковых задач такого типа, для которых эта классификация предназначена.

В начале 1967 г. американские специалисты Р.Фримен и П.Атертон провели интересный эксперимент по использованию УДК в "дескрипторном" режиме. Для эксперимента был использован массив примерно из 250 документов по океанографии, заиндексированных по УДК. Поисковая система была реализована на ЭЦВМ типа ИБМ-1401. Результаты этого эксперимента позволили Р.Фримену и П.Атертон сделать следующие общие выводы:

"Первое и самое очевидное: нет более никаких сомнений, что УДК может быть использована в механизированной системе. Но нужно найти ответ на более трудный вопрос о том, следует ли ее использовать [в таком режиме] /8, с.149/. И далее: "Наш второй вывод ... заключается в следующем: УДК в том виде, в каком она существует в настоящее время, вероятно, не может функционировать в механизированной системе столь же эффективно, как язык индексирования, специально предназначенный для машинной обработки; но не существует никаких препятствий для успешного использования УДК либо в режиме "пакетной" обработки,либо в режиме взаимодействия между человеком и машиной" / 8, с.150/.

Мы в целом присоединяемся к этим выводам, но хотим в заключение подчеркнуть, что

- УДК, созданная в результате интеллектуальных усилий многих ученых и специалистов в течение 70 последних лет, имеет огромную общенаучную ценность;
- УДК нужно всемерно ограждать от вторжений реформаторов, пытающихся тем или иным способом "усовершенствовать" ее, но на самом деле лишь разрушающих эту классификацию;
- УДК таит в себе еще больший запас неиспользованных возможностей для повышения ее эффективности;
- повышение эффективности УДК должно осуществляться главным образом за счет дальнейшего развития заложенных в УДК фасетных принципов, за счет разработки более совершенной методики использования УДК, а также за счет организации быстрого переиздания таблиц УДК.

Цитированная литература

I. Друк Т.М. Использование УДК в режиме координатного индексирования. Сб. "Научно-техническая информация", I968, сер.2, № II, 22-27

2. Курбаков К.И. О принципе "зеркальности" систем классификации информации. (Доклад на 2-м Всесоюзном симпозиуме по научному прогнозированию. 22-26 января I968 г. Москва).

3. Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики.2-е изд., переработ. и допол.М., "Наука",I968, стр.342-344

4. Рассел Б. Человеческое познание. Его сфера и границы. М., Изд-во иностр.лит., I957, стр.277.

5. Шамурин Е.И. Очерки по истории библиотечно-библиографической классификации. Т.П. М., Изд-во Всесоюзной Книжной Палаты, I959, стр.I86-I9I

6. Campbell D.J. Thesaurus of engineering terms: a list of engineering terms and their relationships for use in vocabulary control in indexing and retrieving engineering information. New York, Engineers Joint Council, 1964 [Рецензия]. "Journal of Documentation ", 1965, 21, N 1, 137

7. Costello J.C., Jr. Uniterm indexing principles, problems and solutions. "American Documentation", 1961, 12, N 1, 21

8. Freeman R.R., Atherton P. File organization and search strategy using UDC in mechanized reference retrieval systems. В кн.: K.Samuelson (ed.).Mechanized information storage, retrieval and dissemination.Proceedings of the F.I.D. /I.F.I.P. Joint conference. Rome, June 14-17, 1967). Amsterdam, North-Holland Publishing Co., 1968, pp. 122-152

9. Mooers C.N. Zatocoding applied to mechanical organization of knowledge. "American Documentation", I95I, 2,#I

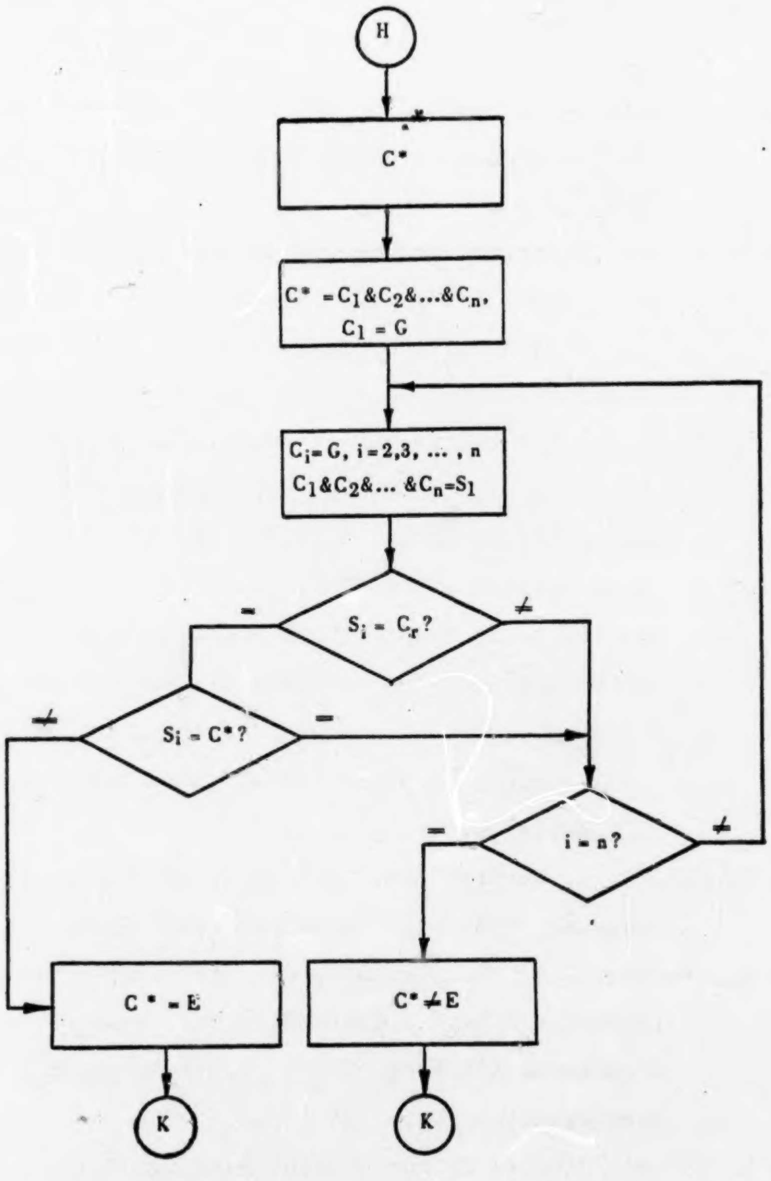
I0. Mooers C. The indexing language of an information retrieval system. В кн. : Simonton W. (ed.) . Information retrieval today. Papers presented at the institute conducted by the Library school and the Center for continuation study, University of Minnesota, Sept. I9-22, I962. Minneapolis, Minn., University of Minnesota, I963, pp. 2I-36

II. Rigby M. Mechanization of the UDC. Final report on pilot project to further explore possibilities for mechanization of Universal Decimal Classification (UDC) schedules. Washington, American Meteorological Society, Meteorological and Geoastrophysical Abstracts, I963, p. 4.

I2. Taube M.,Gull C.D., Wachtel I.S. Unit terms in coordinate indexing. "American Documentation", I952, 3,#4

I3. Taube M., Wachtel I.S. The logical structure of coordinate indexing. В кн.: Taube M. et al. Studies in coordinate indexing. В Vol. I. [Washington], Documentation, Inc. I963, pp. 73-76

I4. Taube M. et al. Studies in coordinate indexing. Vol. II. [Washington], Documentation, Inc. , I954, pp.72-III


Н
C*
C* = C$_1$&C$_2$&...&C$_n$,
C$_1$ = G
C$_i$= G, i = 2,3, ... , n
C$_1$&C$_2$&... &C$_n$=S$_1$
=
≠
S$_i$ = C$_r$?
≠
=
S$_i$ = C*?
=
i = n?
≠
C* = E
C* ≠ E
К
К

Условные обозначения

C^* – анализируемое исходное понятие (класс)

$C_1, C_2, \ldots, C_n$ – понятия (классы), на которые разделяется C^*

C_r – реально существующее понятие (класс)

E – элементарный класс

G – родовое (опорное) понятие в логическом полиноме $C_1 \& C_2 \& \ldots \& C_n$

S_i – логическая структура

Рис. 1. Схема процедуры проверки элементарности класса.

THE UDC
AND AN IDEAL INDEXING LANGUAGE

I.DAHLBERG
Deutsche Gesellschaft für Dokumentation,Frankfurt/Main,Bundesrepublik Deutschland

0. Introduction

The title of the Symposium implies a lot: the UDC which is usually meant to be a classification system is, by the mere word "other", fixed with the purpose of an indexing language. I would not deny such a functional aspect of any classification system, thesaurus or descriptor list, however, I would like to stress, that it should be understood that

> an indexing language is a language with any degree of internal organization useable for indication (hence index) of subject content of documents or data,

whereas, for my opinion, the UDC is not only meant to be an indexing language, but also an arrangement language, a classing and grouping language, a relating language, and perhaps more.

Therefore, any comparison of the UDC with other indexing languages which are not classification systems, like the UDC could not do justice to its other functions than indexing.

With the topic, which I choose, "UDC and an ideal indexing language" I wanted to underline it very purposely that the UDC in its present form is very far from being an ideal indexing language and that we should try to find the ideal indexing language rather than to try to revise the UDC partially or to a great deal.

An ideal UDC could be elaborated if we apply the cognitions which we will gain by comparing the UDC as it is with what it could be and thus the comparison will be on the same level regarding formal and material aspects and regarding the purpose of the classification system; the outcome of this comparison should be nothing else but a more adequate universal classification system.

In this paper I will.however, not as yet outline materially in detail, how an ideal, entirely new universal classification system might look like - this will be tried in a later paper - but I will show in

Part 1 why we need a universal classification as indexing language, in particular

- why a universal classification system is the adequate tool for any indexing work,
- which purposes it should serve and
- what its principles should be.

Part 2 takes a look at the present UDC and states its failures the avoidance of which leads us to

Part 3, where we develop the consequent requirements for a future universal classification system and outline in short its characteristics.

1. The necessity of a universal classification system (U.C.)

1.1 Reasons for the necessary availability of a U.C.

With the rise of the special libraries and the establishment of documentation centers in specialized fields of knowledge more and more specialized classification systems and thesauri have been developed to an extent that the exchange of knowledge was hampered encreasingly and that it became almost generally accepted that a U.C. was an impossibility and that a universal order of knowledge could never more be accomplished. The following reasons as well as the listed purposes and the principles available which all support the idea of a U.C. should encourage a new and perhaps international attempt to strive towards a new conception for a U.C.

1.1.1 A U.C. as a sine-qua-non for universal collections

It goes almost without saying that for a universal library collection or for universal bibliographies, like national bibliographies there is just no better tool than a U.C. for reasonable arrangement and classification.

1.1.2 A U.C. as a valuable tool for general orientation

A U.C., comprising all subject fields and showing their internal relations to each other is equally necessary for educational and encyclopaedic purposes and provides for general orientation that helps to understand Man's environment and his social and intellectual situation.

1.1.3 A U.C. as adequate tool for indexing in special fields

It has been stressed for many times that with a U.C. one could not adequately index in very specific subject fields since a universal scheme is constructed from a universal point of view which necessarily must tear apart the subjects that "belong together" as seen from the point of view of special subject fields.

This, however, is a fallacy, for every field of knowledge - especially in the expanding and integrating world of today - is more or less deeply connected with other fields, neighbourhood fields, supporting fields, application fields and needs the descriptors of these fields in addition to the ones of its own field. Also, every field participates more or less in general categories like space and time, and often needs descriptors for general kinds of objects, materials, products, etc. Often the percentage of the specific descriptors of a field is smaller, compared with the ones from the "outside" fields so that it can be concluded that if the "outside" was a standardized universal system, those developing a special schedule or specific thesaurus would save a lot of effort if they could use the descriptors of this "outside" system.

Also it has been said that with a U.C. one could not index specifically and intensely enough those concepts that only occur in special fields. This of course depends on the usefulness of the structure of a U.C. system and its capacity to include also the most specific terms of special fields. It is indeed hard through a monohierarchical structure with direct subdivisions down to 9 or 12 hierarchical levels to add more concepts and still be reasonable and understandable. But if a U.C. is built with a polyhierarchical structure it should not be impossible to include all the necessary concepts and their combinations from special fields of knowledge.

1.1.4 A U.C. as a universal concept language

Another aspect is the possibility of a U.C. to provide by its notation the intermediary language for the concepts contained therein so that it serves as an international language understandable by every one who can grasp the concepts. The objection brought forward by Wersig (1) that such a universal con-

cept language would be impossible to construct since this presupposes a universality of concepts and universal understanding of the relations between concepts which might have been fixed by the understanding of just one cultural group. But this is a fallacy too since the subsumptive relations establishing a hierarchical classification system are based on logical foundations and can be apprehended just as the concepts thus related by every human being gifted with reason. Only all other relations of the associative kind and the relation of equivalence may be recognized by perhaps only a limited number of persons. But these relations are beyond the necessary structure of a classification scheme and are therefore optative.
Thus such a universal concept language could serve as a tool for educational reasons as well as being a master language for other natural languages.

1.1.5 A U.C. for international exchange

At a time when international exchange of anything, especially however of information becomes an absolute necessity for cooperative documentation it of course, is important to have an international language, a common code, for the semantic designation of the exchanged things or information on subjects. The elaboration of such an internationally useable U.C. should therefore be of paramount concern for all responsible in the field.

1.2 Purposes of a U.C., or what else could be covered by it

So far most of the universal classification systems, like LC, DDC, UDC, Bliss, Colon have been designed to classify documents, mostly books; only UDC was meant to also serve for the classification of journal articles and patents. But there are a lot of other purposes for which universal schemes could be used; E.Wåhlin (2) lists and explains the following other application schemes for classification systems:

1 registration systems (archives)
2 divisions of industry
3 divisions of laws and statutes
4 divisions of occupations
5 classification of articles
    - acquisition of goods
    - customs duty
    - **(international)** statistics
    - catalogues (e.g.of department stores)

6 schemes for standardization

7 schemes for patents

8 industrial classifications
  - for materials
  - for products
  - for technical equipment

9 schemes for cost accounting

10 schemes for education, curricula

11 systems for the editing of systematically arranged compendia, manuals, reference works

12 systems for systematically arranged dictionaries, linguistic thesauri

13 taxonomies for natural objects, like plants, animals

13 museum systems

There is no question that the purposes and items of all of these kinds of used and necessary schemes could not also be found or **assembled** in one comprehensive system; the only provision that must be foreseen is the possibility that such a U.C. have a specific category for objects, the descriptions of which are not only used in a meta-language sense, meaning information about a thing, but also in a demonstrative sense, e.g. as labels of objects.
The purposes of a U.C. therefore could also be universal.

1.3 The principles of a U.C.

There are a number of excellent principles that are applied by the UDC concerning its form, rules and organization which should be recognized as superior to any other existing U.C. In any new attempt the following principles of the UDC should therefore be applied:

1) the decimal notation by which the conceptual system and the hierarchical levels of the classified subjects is expressed,
2) the fact that the main classes of the system are based on subject fields,
3) the separate availability in the paradigma of the system of terms from general categories for combinative indexing,
4) the possibility to apply socalled special auxiliaries by interrupted subdivision,
5) the fact that there exists a collection of rules for the elaboration of the schedules, and
6) an international agency behind the system that takes care

of the elaboration, introduction, maintenance, updating, control, application and its translation into the languages of the world's countries.

If only these principles had been consistently applied to the UDC and if only its rules were based on a sound classification theory and if the organization was such as to guarantee a reasonable development the UDC could have been already the ideal U.C., which everybody needs so badly. But since we mainly learn by experience and failures and since the present state of the UDC leads us to recognize the failures that have been made, it is only natural to strive for a more adequate U.C. The following chapter will show explicitly why we came to these conclusions.

2. The failures of the present UDC

2.1 The structure of the UDC

The structure of the UDC is defined by

- its breakdown of the universe of knowledge into basic classes
- the relations used in the subdivisions of its classes (subject fields, objects, etc.)
- the provision of general and special auxiliaries

The UDC is monochierarchic insofar as the application of direct subdivision in its main tables is concerned, it is partly polychierarchic with regard to the availability of a number of general and special categories.

2.1.1 The basic classes

The universe of knowledge is subdivided in the UDC by the following 10 socalled basic classes:

0 Generalities
1 Philosophy
2 Religion. Theology
3 Social sciences
4 free at the time
5 Mathematics. Natural sciences
6 Applied sciences. Medicine. Technology
7 The arts. Architecture. Photography. Music. Entertainment. Sport
8 Literature. Philology. Linguistics
9 Geography. Biography. History

These basic classes are partly summarizations of subject fields - we shall call these "subject areas ." - like 3 Social sciences or 5 Natural sciences (which latter actually is a mixture of a discipline and a subject area); partly they are subject fields (disciplines), like 1 Philosophy and 2 Religion. Theology; partly they are mixtures of subject fields and subfields, like 7, 8 and 9 and partly generalizations from some other aspect, like 0 Generalities and 6 Applied Sciences. Since these basic classes constitute already from the very beginning a heterogeneously structured material is is only obvious that many of the inconsistencies in the subdivisions of the following levels are due to these very first inadequate forming of classes.

2.1.2 The subdivisions of the basic classes

The two basic classes that are summarizations of subject fields like UDC 3 and 5 can relatively reasonably be subdivided into subject fields, like

3 Social sciences
30 Sociology
31 Statistics
32 Politics
33 Political economy
34 Jurisprudence.Law.Legislation
35 Public administration
36 Social relief and welfare
37 Education
38 Trade.Commerce
39 Ethnography.Folklore

5 Mathematics.Natural sciences
51 Mathematics
52 Astronomy. Geodesy
53 Physics
54 Chemistry. Crystallography. Mineralogy
55 Geology. Meteorology
56 Palaeontology
57 Biology. Anthropology
58 Botany
59 Zoology

The subdivision of a discipline, however, yields subfields, subsubfields, objects, aspects and aspects of subfields, e.g.

2 Religion. Theology	Discipline
21 Natural theology	Subfield
22 Holy scripture.Bible	Object
23 Dogmatic theology	Subsubfield
24 Practical theology	Subfield
25 Pastoral theology	Subfield
26 The Christian Church	Administration Aspect
27 General History of the Christian Church	History aspect of subields
28 Christian Churches and Sects	Objects (Organisation aspect)
29 Non-Christian religions and cults. Mythologies	Objects

Still more "mixed" is the subdivision of those basic classes, which I called "mixtures" as e.g. class 9:

9	Geography.Biography.History	mixture of discipl.and object
91	Geography.Exploration and travel	discipline
92	Biography.Famous lives	object (kind of document)
930	Historical sciences.Collateral and auxiliary sciences	discipline and subfields
940	Mediaeval and modern history	two subfields
950	History of Asia	precombinations
960	History of Africa	of
970	History of North-America	discipline
980	History of South America	and
990	History of oceania, of Australia, of the Polar regions	geographical entities

It must be mentioned that within the subdivisions of **91** there are precombinations with continents and countries just as within history (93/99). These precombinations still have been retained from the old times of Dewey, even the three digit numbers have not been deleted. It must also be mentioned that the old history - also with precombinations with countries is placed under 93; the European history will be found under 940.

2.1.3 The direct subdivision of subject fields

When investigating the relationships occurring in the subdivision of subject fields we find a great number of different methods. It can happen that a field is divided by a number of different facets, like e.g. 02:

o2 Libraries. Librairianship
021 Function,utility,creation,publicity,development, upkeep of libraries
022 Library site and premises
023 Management services. Staff
024 Use of the library. Regulations
025 Library administration, methods
026 Special libraries
027 General libraries
028 Reading. Help and advice for readers
029 free

Or, a field is divided by a mixture of facets and subfields, like e.g. 32:

32	Political science. Politics and current affairs	
321	Forms of political organization. States as political powers	Facet
322	The state and the churches	Facet
323	Internal politics	Subfield
324	Elections. Campaigns, plebiscites, referenda	Facet
325	Migration. Colonization. Colonial policy	Facet
326	Slavery. serfdom. Emancipation	Facet
327	International relations. Foreign policy	Subfield
328	Parliaments. Legislative assemblies	Facet
329	Political parties and movements	Facet

Fields are also subdivided by precombinations with application fields and aspects, like e.g. 368:

368 Insurance

368.1 Insurance of things in general
    .2 Transport insurance
    .3 Life insurance. Personal insurances
    .4 Social insurance. Workmen's insurance. Social security in general
    .5 Agricultural insurance
    .6 free
    .7 free
    .8 Various insurances
    .9 free

The subdivisions are not mutually exclusive: insurance of things against damage or loss can be found under .1, .2 and .5, insurance of people's lives under .3 and .4, insurance against bad chances under .8. Here obviously a subdivision according to the subject/property kind was intended, thus we should only get kinds of insurances. However, under 368.41 we find facets **again, although this special field provides in addition 121 special auxilaries for the specific facets.**

Relative consistency in the application of genus-species relations was found in botany and zoology, whole/part relationship should have been applied e.g. in subdividing the human body, but here the different organs have been subdivided several times according to several points of view. We find the nose under "respiratory system" - 611.21, under "sense organs" - 611.86 and "smell" again under 612.86 and we find it with all the different parts of a nose under 616.211 - diseases of the nose.

If we now examine the rules for subdivision of classes (3) and subclasses, we read (rule 2.21):

> "Each class and subclass is divided according to differentiating characteristics (principles of division) applied successivel y, into as many groups or 'facets' as the subject requires, and the facets are enumerated in the most usual and helpful order".

Now, if the class is a subject field to which we apply this rule, we might, by subdivision according to "differentiating characteristics" always only get again fields, each with a specific characteristic. It is the subsumptive subject/property or the whole/ part relation - if a number of subfields make one field- that is relevant here. But by division according to differentiating characteristics we can never get in a field to the specific facets, which are objects, methods, aspects of objects and methods, etc. Thus rule 2.21 only holds, if a field is subdivided into subfields. There is, however, another rule (2.41)which could perhaps be applied for the direct subdivision of fields into facets, it reads:

> "Facets and features are represented as direct divisions of a main UDC number, when they are of primary importance and peculiar to the subject concerned".

There is, though, no indication about what is meant by "primary importance" and "peculiarity to the subject concerned" and if we examine the schedules, we find e.g. the facet "administration" in a main number (result of direct subdivision) and in a special auxiliary (result of interrupted subdivision),we find all over the method and procedure facet in a direct and in interrupted subdivision. There seems to be no common understanding whatsoever of "primary importance" among those, who built the system and the exemption from rule 2.21 in rule 2.41 leads only to irritation and confusion.

By consequence the UDC schedules cannot be called a systematic classification since the "system-concept, the application of subsumptive relations for the presentation of the hierarchies, lacks consistent use all over and the recent remark of E.Wüster (4) that the UDC contains many pseudohierarchies can only be called a very faint-hearted blame.

2.1.4 The "common" and the "special auxiliaries"

In 1.3 we stated that the availability of general categories and special auxiliaries for concept combination and subdivision of subject fields are excellent features for a U.C.. So far, however, the general categories in the UDC, like concepts for spatial relations, names of countries and subdivisions, descriptors for time, form and language can only be used as "common auxiliaries"; the relevant rule (2.42) reads:

> "Facets and features are represented as -
> common auxiliaries, when of recurrent or subsidiary interest but generally applicable to all or many subjects or areas"

The general practice regarding the space and time category is, however, different from this rule. These are applied for the description of a complex subject just as in Ranganathan's Colon Classification. Concepts and names from the space category are often used as main numbers. The form auxiliary, indicating the kind of document in which some subject is presented, is however, an "auxiliary" of a special kind, since it is not only used for a meta-language information-about-a-thing but also for labelling something as a journal, a dictionary etc.in relation to a subject field. But if such a document (an object) is related to all the subject fields, then the tables say, a main number should be used denoting the object! The other object facets, like persons -05 or the new -03 proposal for materials cannot be used in such a labelling manner, indicating an object. According to rule 2.11 the UDC is only meant "for classifying any kind of document concerning any part of human knowledge..." - it was not meant to classify objects that are not documents! Thus everybody labelling a screw with a UDC-number „misuses" the UDC.

Special auxiliaries can be used in the UDC according to rule 2.42 "when of recurrent or subsidiary interest for limited subject(s) or areas of knowledge". A count of the occurrences of such special auxiliaries in the UDC medium edition produced however a number of 386 places of different lengths and it was not made obvious why they were used just at these places and not somewhere else; also a comparison between some of these places made in (5) showed that they were not developed according to a common plan, each array differing in one field from the other.

It must be concluded that 1) it is apparent for everyone that the structure of the UDC schedules have not been developed under the guidance of a general theory or rule for the subdivision of the fields, 2) that after the existence of the rules, 1961, 1968, the inadequacies in the schedules have not been deleted and 3) - a consequence of 1) and 2) - that apparently the rules are too imprecise to allow for the construction of a logical system.

2.2 Concept combinations, general categories and references

2.2.1 General remarks

Since we have already seen that the classes in the UDC have not been derived by reproducible subsumptive relations, the correctness of the contents of the classes in the different levels of the hierarchy must be questioned. Under these circumstances the application of generic searches becomes a travel into adventureland. The notations, which could not only represent concepts but a concept system turn into the meaning of pure codes. Thus the "system" of the UDC, its meant-to-be hierarchic notation, representing a hierachical system of concepts, looses its characteristicum specificum, its being a system, and becomes finally just a collection of codes.

2.2.2 Precombination of concepts

At many places of the UDC we find as subdivisions of classes concept combinations which are not only combinations due to their hierarchical place but in addition also combinations with other characteristics. The following example shows what is meant:

624.9 Superstructures

624.91 Large-span superstructures, roofs
.92 Solid masonry structures
.93 Timber and composite (bay) structures
.94 Structures of steel. Skeleton structures
.951 Supporting superstructures: platforms, rigs, scaffold frames, etc.
.953 Bulk goods containers, holders, silos
.96 free
.97 Towers. Masts. Pylons
.98 free
.99 free

It can easily be seen that in some of these numbers the material (masonry, timber, steel) forms already the characteristic of a superstructure. In the introduction to this subdivision it can be read: "for materials use .01 from 624.01"; also the material could be indicated by the new proposal from a subdivision of -03. Thus, if we should index a skeleton structure that is not of steel, how would we get the steel out of 624.94 in order to index properly the material of this skeleton structure? The UDC is full of such cases. They probably grew out of the fact, that the special auxiliaries were developed after the precombined subdivisions had existed and nobody had thought of deleting or changing then the notations with the precombinations. I wonder whether the introduction of the material facet -03 will have any influence on all the notations with precombinations with a material in the UDC.

2.2.3 Investigation of occurrencies of general categories in the main tables of the UDC

Although it is claimed that the common auxiliaries provide the necessary general facets and points of view and thus the whole UDC would be a sort of a faceted classification, there exists still a great number of concepts of general categories in the main tables. De Grolier (6) found, that e.g. concepts relating to space occur (in 1962) at 19 places beyond the (1/9) notations, the category "time" was found at 14 other places, questions relating to form and shape exist at 45 and problems of measurement at 78 places. The recognition of these latter two general categories should already in 1962 have lead to a consequent asking for a special group of general auxiliaries for them, but nothing happened.

2.2.4 The see-also references in the UDC

Scanning the tables for such references of the see-also kind renders the impression as if it was tried very diligently to include all the necessary see-also references. But when using them the dissatisfaction with the system can only encrease. The following example will prove this:
If we look for "interpreter" we find a very first place at activities of corporate bodies: 06.053.56 (1) and there a reference to 355.343.18 - Interpreter in military service (2),
356.362 - Interpreter in the administration of the military service (3)
380.835 - Interpreter in tourism (4) but there is no reference to 651.926.08 (5) and none to 82.03 (6). From the notations(2) - (6) there leads absolutely no reference back to any of the previous places.
This is no exception. It can be found dozens of times. The user and indexer is just lost in such a jungle of placings under so many different viewpoints, not to speak of the job of the searcher.
This example also proves that the introduction of the common auxiliary of persons -05 has had no influence on the contents

of the main tables. With the possibility of fixing an "interpreter generally" by adding the -o5 to 651.926 all the other places could have been deleted or their special aspect could have been made explicit by colon combination. This would have cleared up the situation for the user immensely. But apparently this was not only forgotten but could not be done because it would have been an act against the rules, since rule 2.44 states that "the order 2.41 to 2.43 (that is: direct division of a main number, special auxiliaries, common auxiliaries, colon-related subdivisions) (foregoing) is the preferred order of choice in deciding between two or more possible ways of classifying or indexing".

2.2.5 The colon-related subdivisions

According to rule 2.43: "Facets and features are represented as colon-related subdivisions, borrowed from other subject-fields scheduled in the UDC tables" one should think that the UDC is a complete faceted classification which builds up its complex, polyhierarchic concepts by colon-combinations of two or more notations from the main schedules. This extremely practical feature was however not made the first principle of subdivision (as we have seen in 2.2.4) but the last one and secondly it was apparently discovered relatively late so that it could not be applied any more where a precombination already had been taken place. Thus the tables are full of unnecessary precombinations as could be seen in many of the examples given and as can be proved many times more. Wåhlin(in 2,p.3:5) shows that the element selenium occurs no less than seven times, sulfur can be found in the Abridged Edition 14 times (together with its compounds), plants are placed in botany just as in agriculture etc.

It would lead too far to try to reveal all these inconsistencies in detail. But from these few hints it should be obvious that the quality of the relations in the UDC proves once more that the system cannot even be revised any more, it must be rebuilt entirely.

2.3 Quantitative aspects

2.3.1 Quantity of concepts in the UDC vs such in thesauri

Comparisons of existing thesauri and the UDC regarding the amount of concepts included showed that the UDC which claims to be a universal system and applicable for special fields too, fails to provide the same amount of concepts that special subject thesauri give in specific fields. The investigation of Vickery and Slater (7) which was also interpreted by H.Arntz (8) had been conducted in the field of nuclear energy - for which a special UDC edition exists - and also in other fields of natural science. In the first case the Euratom thesaurus used 247 concepts in a specific field, the UDC however only 181 - a 25% less; in the second case the Thesaurus TEST was compared with a number of universal classifications among which the UDC. Here for the 49 concepts in TEST only 29 could be found in the UDC. The other universal classifications yielded worse results.*)

2.3.2 Quantities of concepts in UDC vs distribution of subject fields and literature produced in these

In a Czechoslovakian investigation (9) the following numbers are given for the distribution of UDC numbers in the different basic classes:

(on the basis of the UDC full editinn)

o	1.549
1	1.732
2	2.289
3	13.397
5	23.849
6	67.274
7	4.482
8	923
9	1.230

That means that in the classes 3,5 and 6 of the UDC full edition more than 83% of all UDC numbers are concentrated (with the exception of the common auxiliaries of which 8.220 exist). The numbers for the distribution of the documents according to the basic classes even yields 89% for the three classes mentioned as will be seen by the following amounts:

0	15.000	6	650.000
1	25.000	7	30.000
2	10.000	8	30.000
3	90.000	9	30.000
5	400.000		

---

*) It must, however, be mentioned here, that recent investigations by Wellisch and de Vries (16) as well as by Čhman and Olivecrona (17) have shown that perhaps due to the wrong UDC editions used, the results of the Vickery and Slater studies are not as accurate as they could be. Wellisch found that between 80-90% of the TEST terms could be expressed by UDC notations.

This uneven distribution is of course due to the fact that the creator of the decimal classification had only thought of devising a scheme for a small liberal arts college and not for a highly industrialized society. That his scheme is "abused" by his successors was not Dewey's fault. This results, however, also in extremely long numbers for discriminating concepts in technology and in the unavoidable combinations of such long numbers for the forming of complex concepts. The relevant rule, however, demands: (2.33):

> "The essential purpose of the notation is to permit easy and rapid retrieval of classified or indexed documents. Consequently it should al ways be as simple and concise as possible: very long numbers are a source of error and inefficiency"

Those facing these unbalanced quantities of numbers and inapplicability of the rule 2.33 in the three important basic classes must be asked how they want to win a competition with other indexing systems under all these circumstances?

2.3.3 Rules relating to quantitative aspects in the UDC

There are two rules concerning the existence of duplicate placings in the UDC. Rule 1.32 says:

> "Duplicate placings of the same topic should be avoided: before proposing a new number, a careful check should be made by the proposer, that the concept is not already included somewhere in the UDC"

Rule 1.41 says:

> "Future development of the UDC will be facilitated, if existing duplicate and superfluous numbers are rendered vacant or 'free' by cancellation".

When checking the UDC even in the most recent numbers e.g. P.71-9 it seems however, that these rules have been completely forgotten. New numbers are added and all the ones that could be deleted,are retained. And this is not done because it was perhaps forgotter since in this very case I have myself pointed out to the resposible persons that the superfluous numbers could be eliminated. There is just no excuse for such an attitude. The argument, that the different aspects under which objects could be seen justify their placing in numerous hierarchies is not a solid reason, since for the searcher it is not always obvious, under which aspect an object necessarily had been classified. One should eather proceed in an opposite manner and make the objects the main classes (in the object areaus) so that thery are the constants and the aspects under which they can be seen or treated could be made the facets,the variables.

2.4 Methodology of the use of UDC numbers

(By methodology is meant the use of UDC-numbers as indexing language for the forming of propositions; it is not meant the filing of UDC-numbers.)

2.4.1 Paradigmatic organization as basis and source for syntagmatic formulation of propositions

The function of a classification system must be seen in the supply of elements from a structure for the forming of informative propositions. A system (paradigma) is constructed the more flexible for a presentation of combinations of subject content, if its elements have been defined in a unique and limiting sense, that is: the less precombinations of subjects are contained in a system the better the basis for postcoordinate propositions.

The great number of precombinations in the tables of the UDC and a consequent lack of single concepts render difficult the use of UDC numbers for informative indexing.

2.4.2 Semantic relationships

When forming a proposition by use of several UDC numbers, the relationship yielded by coloning is generally not semantially qualified. In order to give a meaning to such a relationship some authors have made proposals for the use of specific numberor letter-codes, thus D.Kérvégant (10) and J.Perreault(11). The latter proposed a structured system of 108 realtors; these were adopted for provisional use with UDC numbers (12).Such relators can, however, not be used for concepts which already contain syntactical precombinations. e.g.

338.89 measures against monopols
343.92 physical influences on criminal actions.

Thus the same conclusion that is true for 2.4.1 can be applied here too: as long as the concepts in the UDC are not singled out according to their unique subsumptive or faceted relationship the use of relationship indications cannot be used for the improvement of a proposition on a subject.

2.4.3 Sequence of UDC-numbers

The sequence of the UDC numbers in a complex UDC expression has to follow a certain formal order. This was established so that clerks could more easily file and find catalog cards with UDC numbers. However, on account of this formal order it is not always possible to reconstruct the meaning of a complex UDC expression. This again is a reason for loss of information. It should be possible to read from the sequence and arrangement of numbers, possibly with the addition of relators the meaning of the proposition that is included in them.

3. Statements and conclusions

From the preceding critique we shall now derive the following 14 statements and develop from each of these corresponding conclusions which should help us to find guidelines for the development of a better tool for universal indexing.

3.1 Statements concerning the structure

Statement 1: The basic classes of the UDC are not based on summarizations of disciplines (subject areas) but on a mixtum of a) subject areas - like UDC 3 and 5, b) subject fields - like UDC 1 and 2., c) mixtures of subject fields and subfields, like 7,8 and 9 and d) generalizations from some other aspect like 0 and 6.

Conclusion 1: In order to obtain a consistent breakdown of the universe of knowledge into subject fields, it is necessary to form reasonable summarizations of subject fields, "super-classes", which put the "elements", the fields, into a distinct order.

Statement 2: The first subdivision of the basic classes of the UDC is not based on fields alone but on fields, objects, aspects and the like.

Conclusion 2: In order to subdivide a basic class consistently in all areas , it is necessary to subdivide it only by fields, not by certain facets. The fields equally then must be subdivided into subfields, if applicable.

Statement 3: In the UDC the subdivisions of the subject fields are not accomplished by application of a certain rule indicating the contents or array of specific facets. The order and contents of the facets which can be obtained by direct or interrupted subdivision is completely left to the judgment and free will of those, who develop a certain field.

Conclusion 3: In order to develop a consistent U.C. it is necessary to fix rules for the subdivision of fields which provide for interrupted subdivision as soon as the specific facets of a field begin to be treated and by which the facets are arranged in a common order. For this a proposal can be found in (5).

Statement 4: The characteristics of a class in the UDC often do not exclude each other. When synthesizing, e.g., a number of concepts from subdivisions of a class, then such numbers cannot be searched separately.

Thus, whenever a search is directed towards one specific aspect concept of a synthesized everytime information is lost if this specific aspect was not indicated by interrupted subdivision.

Conclusion 4: All the variables of a class (all its variable aspects) must be developed by separable, linkable concepts. The principle of "interrupted subdivision" should be applied in subdividing recurring attributes.

Statement 5: Although in the UDC common auxiliaries are existing for space, time and language subdivisions and are used as main numbers too, some of these have openly and in a hidden way been built into main numbers, e.g. classes 8 and 9.

Conclusion 5: General facets should never be merged with the names of fields. In order to express combinations of fields with items of such general facets colon-combinations of concepts should be established.

Statement 6: The subdivisions of the UDC are in many cases not based on subsumptive relations (genus-species, whole-part, subject-property) but on a mixture of these and specific facets often with leaps in the different steps of the concept-ladder to that socalled pseudo-hierarchies occur.

Conclusion 6: It should be taken care that direct subdivision of a class is always only accomplished by application of a specific kind of subsumptive relationship.

Statement 7: The UDC besides including a lot of precombined concepts in its main tables provides also rules for analytic-synthetical presentation of concepts. Thus, on the one hand, the system is unnecessarily redundant, on the other hand it does not render enough of the single concepts for concept combinations.

Conclusion 7: A newly structured U.C. ought to avoid unnecessary precombinations of concepts; if still such should be deemed helpful, care should be taken by the index that a specific concept then always may only be expressed by an established precombination and not in any other way.

3.2 Statements concerning the contents

Statement 8: Although the UDC claims by application of a number of common auxiliaries to be a faceted classification, it still contains a significant number of concepts from general facets among the socalled main numbers in the main tables.

Conclusion 8: When developing a new U.C., careful separation of concepts from general categories like space, time, form, measurement etc. should take place so that these are not hidden in the schedules but made applicable to all fields of knowledge.

Statement 9: At many places in the UDC disciplines and their practical applications have been put together, as e.g., 2 Theology and Religion, or, 8 Linguistics and Philology. In natural sciences, however, technological applications have been classified apart, as e.g. 54 chemistry and 66 chemical technology.

Conclusion 9: In a new U.C., care should be taken, that a separation of a science and its technological or practical application should not be admitted.

Statement 10: The distribution of UDC numbers in the different main classes is extremely unequal. Less than one third of the basic classes provide 83% of the total amount of concepts.

Conclusion 10: By considering the contents of a subject area care should be taken that the distribution of subject fields and specific concepts somehow complies with the actual production of knowledge in this area

Statement 11: A comparison of UDC concepts with the amount of literature produced in one year not only supports statement 10 by revealing that almost 90% (correctly 89%) of the documents published belogn into only three basic classes of the UDC; it also shows that future development will presumeably enlarge the so far absolutely unbalanced state. Class 6 is growing out of the whole system like the tower of Babel.

Conclusion 11: In developing a new scheme care should be taken to consider existing and future development of knowledge.

Statement 12: In the UDC big classes of objects, like chemical elements, materials, products are listed at a number of places; objects, like plants and animals are listed under the main numbers, persons, however, among the common auxiliaries.

Conclusion 12: It is proposed that in a new scheme the classes for general objects be excluded from the subdivisions of subject fields in order that these can be listed only once and be made constants in a system where all the aspects under which they can be seen are made the variables.

3.3 Statements concerning methodology of use

Statement 13: Semantic relations (like relators) can only be used with thoscUDC-concepts that are not already precombinations with built-in semantic relationships.

Conclusion 13: In order that propositions may be formed from single concepts in syntagmatic organization, it is necessary to define the facets of a given field in such a way that their elements may be used in combinations with relators.

Statement 14: The sequence of the concepts in a complex UDC number is to be followed according to a formal convention, which is useful for filing, it however does not supply a semantic help for the proposition in an indexing "formula".

Conclusion 14: The rules of a new U.C. should provide the possibility to derive the semantic information of a proposition from the sequence or arrangement of the concepts in an indexing formula.

3.4 Summarization of conclusions:

The following findings have only been derived from the above conclusions. They don't as yet draw a complete picture for a future U.C. This should be kept in mind. We do need still more research about the contents of subject fields and their interdependency with other fields in order to be able to develop a sound theory that could support a useful design of a new U.C.

3.4.1 The structure of a U.C.

The universe of knowledge should be subdivided into
- a) subject areas and these into subject fields,
- b) general categories among which we also find a number of object fields, e.g. for materials, plants, animals, persons, societies, products, documents, languages

The subject fields should be subdivided by a standard chain of facets in interrupted subdivision. The object fields could also be subdivided in this way of interrupted subdivision; here the principle of subdivision must always only follow subsumptive relationhips.

3.4.2 Contents of a U.C.

A U.C. should contain all the concepts and names of objects that are necessary for the indication of any subject content or labelling of things.It should list all of these in surveyable tables, indicate the necessary precombinations of terms into combined and/or complex concepts and make them accessible together with the other single concepts and terms through an alphabetical index. In a thesaurus-like manner equivalent terms and concept associations could be included.

3.4.3 Methodology of the use of a U.C.

Ideal indexing could be accomplished if a U.C. provides
- a) an easy to overlook paradigmatic organization where each object occupies just only one place so that concept combinations for the formulation of propositions are facilitated,
- b) a number of classes of semantic relators for the indication of syntactic relationships,
- c) the necessary rules for the semantic sequence of the concepts.

4. Concluding remarks

I have only been able to touch some special weaknesses of the UDC and did not really come to statements which the knowledgeable did not already also know ten years ago. There exists an FID document of 1962 which lists the essential mistakes in the UDC although not as explicitly as it was done here but which proves also that already then a rigorous desire to revise the system rather thoroughly had existed.
It is true that 25 revision committees today try to adapt the UDC to present day knowledge, but how can the result of their work alter the UDC into a more consistent indexing tool if they may only repair here and there some rooms in a medieval castle instead of constructing a modern building with the accomplishments of modern classification theory and -technique? Already in 1962 it was proposed to develop a completely new system, but then it was decided to "investigate whether an equally or similarly good system could not be reached by changes accomplished in the frame of the existing UDC" (13).

Whether this decision was right or wrong I would not dare to judge, but I would like to predict that the status of the present UDC and its revision rules are not of the quality to secure for this system in the future the necessary application in international information systems. Although the present UDC might still be alive for another 10 years, it is almost irresponsible today to teach and recommend the use of the UDC and it is equally dangerous to advise the use of only some starting numbers as "umbrella or roof classification". In the FID publication "Key to information" it is proposed -

> "The best prospect of achieving a reasonable degree of compatibility among them (different indexing systems, I.D.) seems to depend on linking such systems with a general classification like the UDC, covering all disciplines, in combination with a general thesaurus or thesauri" (14)

and the dreamers hope that the UDC "may well become the major 'switch language' in the exchange of information in fully automated networks". A general rule providing perhaps all the UDC numbers denoting a subject field with a link to the descriptors of relevant thesauri would not rescue the UDC at all and would if really applied only confuse the situation more and more. We should rather stop hiding our heads in the sand and stop wasting money for unreasonable revision work but start to learn from our failures, start to develop a theory for a better U.C., search for an organization who would support this - if not the FID then perhaps Unesco - H.Wellisch has well enough recently pointed out FID's principle organizational obstacles (15) - and finally organize the necessary elaboration work according to a sound theory.

Literature:

(1) Wersig,G.: Rette die DK wer kann! - In: Nachr.Dok.22 (1971)Nr.2, S.64-65

(2) Wåhlin,E.: Classification systems. - Stockholm 1971.

(3) International Federation for Documentation: Universal Decimal Classification. Revision and Publication Procedure. - The Hague: FID 1968. 53 p. = FID 429

(4) Wüster,E.: Begriffs- und Themaklassifikationen. - In: Nachr.Dok. 22(1971)Nr.3

(5) Dahlberg,I.: Möglichkeiten einer Neugestaltung der DK. In: Nachr.Dok. 21(1970)Nr.4, S.143-151. In English in: J.Doc. 27(1971)No.1, p.18-36 (Title: Possibilities for a new universal decimal classification.)

(6) Grolier,E.de: A study of general categories applicable to classification and coding in documentation. - Paris: Unesco 1962.

(7) Vickery,B.C.,Slater,M.: Classification in science information. Comparative study on classification. Paris: Unesco 1969. Paper CSI 5.8, 30 Oct.1969.

(8) Arntz,H.: Die DK - eine Vielfacettenklassifikation. In: Nachr.Dok. 21(1970)No.4, p.139-142

(9) Kofnovec,L., Simandl.D.: Comparison of UDC structure with structure of world scientific and special literature. Prague 1970. Working paper, 7 p.

(10) Kervégant,D.: Bull.Bibl.France 4(1959)No.11, p.502-3

(11) Perreault,J.: Towards a theory for UDC. - London: C.Bingley 1969. (here p.119-148)

(12) PE 958. Auxiliary Table I.b Relation - Scheme of Relators. The Hague: FID 1969. 4 p.

(13) Die zukünftige Gestaltung der DK. Empfehlungen des zentralen Klassifikationsausschusses (FID/CCC). Den Haag: FID 1962.

(14) International Federation for Documentation: Key to information. Universal Decimal Classification.- The Hague: FID 1970. 14 p. = FID publ.466

(15) Wellisch,H.: Organisatorische Neuordnung des DK-Systems. In: Nachr.Dok. 22(1971)No.2, p.55-63

(16) Wellisch,H., Vries,I.de: A concordance between UDC and TEST; results of a pilot project. - Paper of Internatl.Symposium, UDC in relation to other indexing languages, Herceg Novi, Yugoslavia, June 28-July 1,1971.

(17) Öhman,E., Olivecrona,C.: Some notational, hierarchic and syntactic problems in connection with concordances between UDC and thesauri. - Paper of Internatl.Symposium, UDC in relation to other indexing languages, Herceg Novi,

PROBLEMS
OF INDEXING AND CLASSIFICATION IN SMALL, DEVELOPING AND MULTILINGUAL COUNTRIES

Ljubomir DULOVIĆ,Milutin GAVRILOV
Yugoslav Center for Technical and Scientific Documentation
Beograd, Jugoslavija

Introduction

The position of SFR Yugoslavia in the world, the degree of its scientific, cultural and economic development, the number of languages spoken by its peoples and nationalities, as well as the number of existing and potential users of information, have induced the authors to investigate the relationship between UDC and other indexing languages. Another relevant factor in this context is the fact that our country "imports" over 99% of world science. If we accept the postulate that the existence and further progress of developing countries directly depend on their ability to make use of world science, it obviously follows that a systematic study and utilization of new world science are essential.

The present symposium has been organized in order to consider the idea of adopting an international indexing language, making possible the universal use of all processed and indexed information. As you may remember, the proposal to hold this symposium was made by the representative of the Yugoslav Centre for Technical and Scientific Documentation at the Second International Symposium "UDC and Mechanized Information Systems", held in Frankfurt/Main in June last year.

Linguistic barriers in the use of primary world literature cause numerous problems and difficulties, in our as well as in other countries. Such problems and difficulties can be avoided in the use of secondary and tertiary literature if a common artificial language, an indexing and classification language, is adopted.

Over 300 years ago, Gottfried Wilhelm Leibnitz, the author of the integral and differential calculus, became aware of that problem of modern world. He pointed out that it was possible to reduce all complex terms and notions of human knowledge to their simplest elements, which could be expressed by mathematical symbols. Since these symbols are understandable to everybody (just like numbers and algebraic and chemical symbols), it would be possible thus to obtain a supranational language,

This issue, which Leibnitz noticed and partly resolved a long time ago, represents in fact the cornerstone of the contemporary information science. The storing of information in the most condensed form possible and its subsequent retrieval can be organized on a world scale, provided that a common world indexing language exists and is used.

Arguments for and against UDC

The work and the studies conducted within the UNESCO project for the UNISIST world information system differ in their assessment of UDC. It was considered at first that UDC was almost certainly unsuitable for UNISIST. A. Wysocki, the UNESCO official in charge of UNISIST, was of the opinion that UDC classification could be used for UNISIST, but with certain modifications.

The latest views of UNESCO concerning UDC are presented in the book "UNISIST - Synopsis of the feasibility study on a World Science Information System" which was published this year. It says about UDC:

"Its further potential has yet to be realized, and both a continuing programme to strengthen UDC and further studies and experiments to test its applicability to retrieval systems are desirable".

Work on UNISIST has been going on since January 1967 but, apart from discussions, no practical results have materialized so far.

According to some estimates, UNISIST should be ready for active use in ten years time. However, the question is: what to do until then, and whether the newly adopted universal indexing language will not have the same shortcomings for which UDC and all other thesauri are today criticised.

At the initiative and proposal of Dr.H.Arntz, a group of three experts (Messrs. A. van der Laan, G.A. Lloyd and J.H. de Wijn) have written a paper on the possibility of using UDC to co-ordinate the other classification systems.

There seems to be no doubt that the further development and application of UDC depend on the extent to which it will meet the requirements made by UNISIST.

In addition to Dr.Arntz, the problem of the reorganization of UDC has been studied by H. Wellisch, J.Dahlberg, G.Wersig and other authors.

Two international symposia on UDC (Kopenhagen 1967 and Frankfurt 1970) have demonstrated that UDC is applicable to automated information systems, like all other indexing languages and thesauri. Automation offers a broader scope to the application of UDC. Machine-readable files of secondary documents indexed in UDC can be easily translated into those languages in which corresponding UDC editions exist.

There does not seem to be much sense in quoting the arguments of many authors and eminent experts in the defense of, or against, the application of UDC as a universal classification system, as this would not help much to clarify the situation.

When speaking about UDC, it should be borne in mind that it is actively used almost throughout the world, and that its applicability is increasingly studied and discussed. However, all suggestions and proposals as to further measures are open to question. That question is evidently as complex and open as was the question what the astronauts would find on the Moon.

For all small nations, developing countries and multilingual countries UDC represents not only a classification system but also an international language, with files available to all users under the same conditions.

Indexing and classification in SFR Yugoslavia

A "Guide through Information and Documentation Units in Yugoslavia" was published in 1967 by the Association of Engineers and Technicians of Slovenia. For each of the 234 information and documentation units (INDOC) eleven data are given in the book. From the statistics provided in that guide it appears that 107 INDOC units (approx.46%) use the Universal Decimal Classification for their primary and secondary documentation. The next largest group of INDOC units (80 units, i.e. approx. 34%) use for this purpose various types of catalogues. The "individual decimal classification" has 22 users among these INDOC units (close to 9.5%), while only 9 units reported the use of thesauri. Among the remaining 16 INDOC units the dispersion is so great that no more than two units apply the same system.

It may be useful to mention at this point that the Yugoslav Centre for Technical and Scientific Documentation in Belgrade carried out a survey in 1968 and found that over 2200 Yugoslav libraries used UDC.

It is evident from these figures that UDC has a considerable number of users in Yugoslavia. Acting on this information, the Yugoslav Centre for Technical and Scientific Documentation has launched a large-scale programme of translating UDC editions into Serbo-Croat, Slovene, Macedonian and Albanian.

It should be also pointed out that Yugoslavia, as a developing country, has particular problems in classification and indexing because of its multilingual composition. To mention just the most important languages spoken in our country: Serbo-Croat, Slovene, Macedonian, Albanian and Hungarian.

Conclusion

We believe that there is little chance that one of the world languages will be chosen as intermediate for secondary documents. The main reasons for this belief are of a political and psychological nature.

The privileged position and advantage of developed countries and of the countries in which world languages are spoken in availing themselves of the world science are obvious The education and development gap between the "big" and the "small", the developed and the underdeveloped is not growing narrower but, on the contrary, wider. The time needed to gain access to new knowledge is becoming relatively longer, and the financial resources needed to use that knowledge are so great that many countries cannot afford them.

Some interesting conclusions are made in the A [illegible]-ICSU report on "Classification in Science Information" ([illegible] document C 70-12, 3rd April 1970). The comparative investigation of several classification systems have revealed that UDC is the least unsuitable of all the investigated systems.

The developing countries expect, and justly demand, that UNESCO should bear in mind while developing the UNISIST world information system that the rate of their growth and the possibility to bridge the scientific and technological gap greatly depends on the utilisation of secondary publications, and on the indexing language that is used in this.

In one of his papers, M. Rigby voices doubts as to the feasibility of evolving such a classification system as would meet the linear and hierarchic as well as the co-ordinate requirements. He goes on to predict that an existing universal scheme would be eventually adopted, and supplemented with a number of natural language or special schemes, none of which would be 100% concordant with the universal classification. He further believes that by such a combination the universal decimal classification would be rendered more flexible, more responsive to new concepts or areas, more idiomatic and less stable.

In the article under the title "UDC number - a useless requisite"?, the author states that "in the field of scientific and technical information UDC is today the most widespread common means of communication in the world".

This is an irrefutable fact. Until other indexing languages are thoroughly investigated, UDC will remain a system which produces maximum effect.

INDEXING AND RETRIEVAL OF RECORDS OF BRITISH PATENT SPECIFICATIONS USING THE UNIVERSAL DECIMAL CLASSIFICATION AND THE BISRA EXPERIMENTAL THESAURUS

Richard HINDSON

British Steel Corporation,Motherwell, Scotland

Origin of project

This project had its origin in the fact that various units, including the information services of the then geographical groups of the British Steel Corporation, the patents group of the British Iron and Steel Research Association (BISRA) and the BSC Head Office patents department were, in 1968/69, all monitoring and/or utilising mainly for information purposes, numerous published British patent specifications and some foreign specifications. It was felt that there was considerable overlap in the specifications selected, disseminated and indexed by each unit. Also, the view had been held by a number of people for some time that it might be useful to organise a centralised computer based system to record, for future retrieval, details of patent specifications relevant to the activities of the British Steel Corporation. Fortunately some experience in the organisation of computer based systems for the handling of details of publications including patent specifications had already been gained by BSC information staff located in Scotland. In particular, a search program for retrieving bibliographical data from a fixed field magnetic tape record had been devised and used experimentally. Thus a situation existed where it was possible to consider the organisation of an experimental project which would provide practical experience of the various factors involved in operating a centralised computer based patents information retrieval system.

Preliminary activities

The immediate need was to quantify the volume of material which such a system would include. Therefore over a period of seven weeks (20th February 1969 - 10th April 1969) the British patent specifications handled by BISRA, the Swinden Laboratories Information Service and our own Information Service were recorded and analysed. During this period 230 British patent specifications were selected by the three organisations. Of the 230, 61 were common to all three organisations; 53 were common to only two of the organisations and the remaining 116, or approximately 50% were unique to one organisation. It was estimated that the total number of specifications handled by all relevant units within the BSC would not exceed 60/70 per week, of which around 10% would be of foreign origin. (It was accepted that the number of foreign specifications selected at that time was only a very small proportion of the number relevant to the activities of the Corporation).

After discussions and consultation had taken place a meeting was held in Scotland to agree upon the design of a suitable system for the handling of the bibliographical details of patent specifications. It was decided that only British specifications would be included since if details of all relevant foreign specifications were included an insupportable burden would be placed both on the monitoring sources and the central facility. Since the search program operated on a fixed field layout this predetermined the design of the item record. Specimen card and paper tape layouts for that part of the record searched by the interrogation program are shown in Illustrations 1 and 2. Each item record includes UDC notation and keywords selected from the BISRA experimental thesaurus.

Abstracts were prepared separately and the relevant paper tape layout is shown in Illustration 3. (No provision exists for the input of abstracts by means of cards).

As paper tape input was, in fact, used throughout the project the full list of tag numbers used is shown in Illustration 4.

4

The BSC owned computer installation originally used for this project was an ICL 1903.

This comprised

1903 16K	Processor
2102	Card Reader
1916	Paper Tape Reader
1971	Magnetic Tape Deck (4 Tapes)
1933	Line Printer
1920	Card Punch
1925	Paper Tape Punch
2802/2	5 Exchangeable Disc Stores

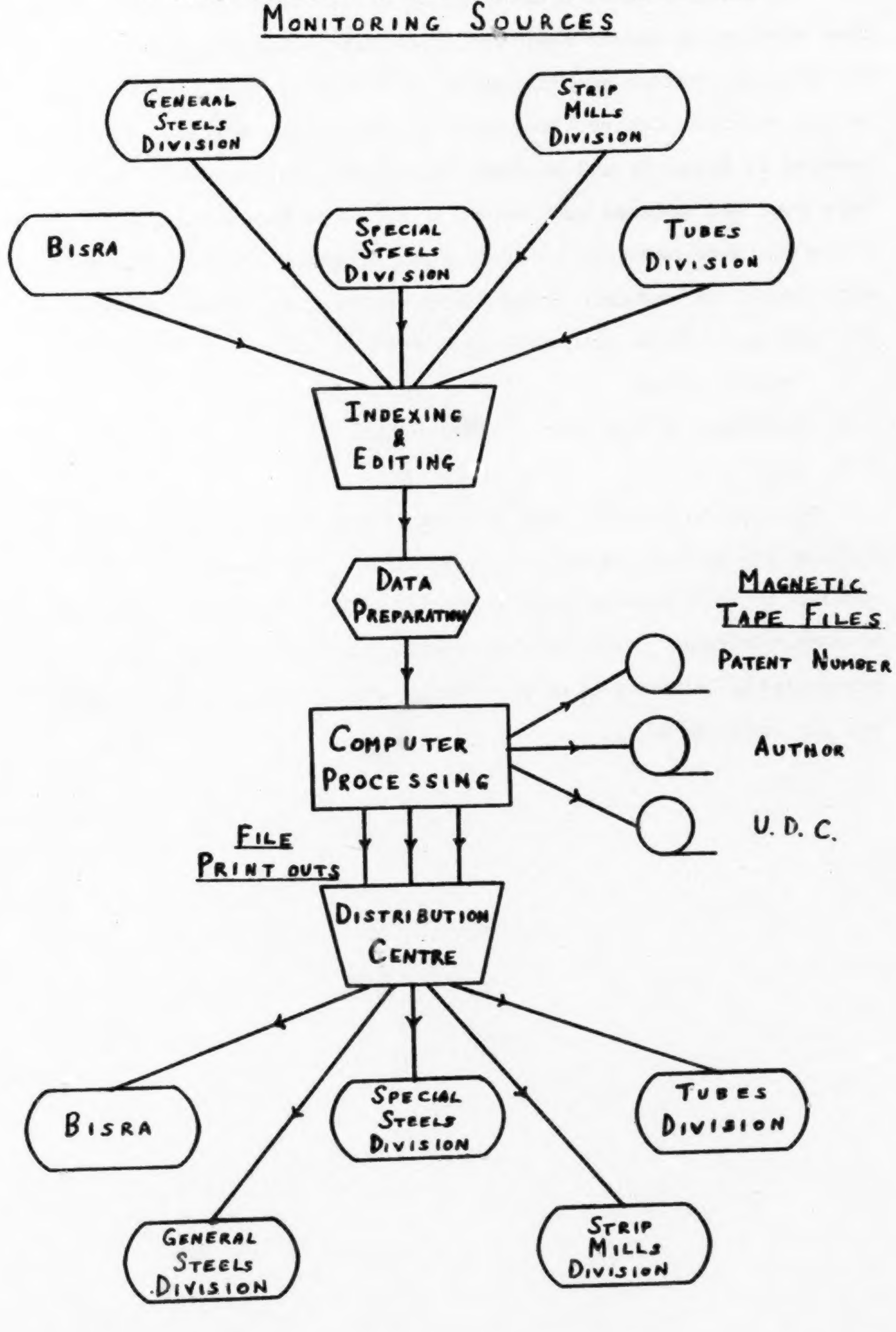
Recently this installation has been replaced by a 1904 A configuration comprising

1904 A 65K	Central Processor
7901	Communications Processor (not yet installed)
-	Line scanning and Termination Units associated with the 7901
2802/2	8 Exchangeable Disc Stores
1972/2	Magnetic Tape Decks (4 Tapes)
1971/2	Magnetic Tape Decks (4 Tapes)
2102	Card Reader
1920	Card Punch
1916	2 Paper Tape Readers
1925	Paper Tape Punch
1933	2 Line Printers

It is intended that on-line facilities will be available.

Clearly an experimental system such as the one envisaged could not be run without staff at the central point. In the event, one information assistant and one clerkess were appointed to handle the editing, indexing and punching.

It was planned that the overall system would operate thus:


MONITORING SOURCES
GENERAL STEELS DIVISION
STRIP MILLS DIVISION
BISRA
SPECIAL STEELS DIVISION
TUBES DIVISION
INDEXING & EDITING
DATA PREPARATION
MAGNETIC TAPE FILES
PATENT NUMBER
AUTHOR
U. D. C.
COMPUTER PROCESSING
FILE PRINT OUTS
DISTRIBUTION CENTRE
BISRA
SPECIAL STEELS DIVISION
TUBES DIVISION
GENERAL STEELS DIVISION
STRIP MILLS DIVISION

Development of data base

The work of constructing the data base began in January 1970. Each monitoring source supplied at regular intervals details of the British patents specifications selected by them. The staff at the Research Centre (Scotland) at Motherwell edited, classified, indexed by keywords and punched the entries for input. The data base was updated each month. From the beginning cumulated printouts were provided so that any monitoring source could have, each month, if desired, a record of all specifications input to the system. These printouts gave entries under

1.	Patent number	(Illustration 5)
2.	Patentee or assignee	(Illustration 6)
3.	UDC	(Illustration 7)

No keyword indexes were provided because no suitable computer program had at that stage been developed. By the end of 1970 details of 2079 British patent specifications had been recorded on magnetic tape. The data base was thus large enough for experimental interrogation on a batch basis, and therefore input was put into obeyance.

Experimental interrogation

After consultation with our technical colleagues 50 experimental questions were constructed. Of these 10 were actually run earlier this year (1971). These 10 questions were:-

1. a) Question in natural language.

New processes for production of steel.

b) UDC Translation.

669.18-932) 669.18-932)	Continuous steelmaking processes
669.141.1) 669.141.13)	Direct process steel
669.181) 669.181.274)	Direct process for production of steel
669.054.7) 669.054.7)	Electroslag refining

c) Keyword Translation.

Steelmaking and continuous process
Spray refining
Electroslag process
Direct reduction
Basset process
Cyclosteel process
Dored process
D-LM process
Echeverria process
Fior process
Flame smelting
Fluid bed reduction
H-iron process
Hoganas process
Hyl process
Krupp-Renn process
Nu-iron process
Purofer process
R-N process
Strategic Udy
S-L process
Tysland-Hole
Wiberg process

2. a) Question in natural language.

Oxygen degassing of molten steel in ladle.

b) UDC Translation.

669.046.554) deoxidation in ladle
669.046.554)

669.183.218.554) OH process,
669.183.218.554.9) deoxidation in ladle.
669.184.235.54) BO Steelmaking
669.184.235.549) deoxidation in ladle

c) Keyword Translation.

Deoxygenation & ladles.

3. a) Question in natural language.

Blast furnace tuyeres design.

b) UDC Translation.

669.162.221) Design of blast furnace tuyeres
669.162.221.8)

c) Keyword Translation.

Tuyeres & blast furnaces & design.

4. a) Question in natural language.

Utilisation of blast furnace slag.

b) UDC Translation.

669.054.82) Utilisation & 669.162.275.2.) Blast furnace
669.054.82) of slags 669.162.275.2) slags

c) Keyword Translation.

Utilisation & blast furnaces & slags.

5. a) Question in natural language.

Improved rails.

b) UDC Translation.

669.14-424) Rail sections.
669.14-424)

621.771.262) Rail section mills
621.771.262)

c) Keyword Translation.

Rails & improvements

6. a) Question in natural language.

External desulphurisation of iron.

b) UDC Translation.

669.046.546) desulphurisation
669.046.546.24)

c) Keyword Translation.

Desulphurising

7. a) Question in natural language.

Improvements in transportation of hot metal.

b) UDC Translation.

621.746.22) transport of molten metals to mould
621.746.224)

c) Keyword Translation.

Transportation & liquid metals
Bulk transportation and liquid metals.

8. a) Question in natural language.

Engineering improvements in heavy plate mills.

b) UDC Translation.

621.771.23)	Heavy plate	&	621.771.06)	Design &
621.771.23-413)	mills		621.771.06)	construction of mills

c) Keyword Translation.

Plate & rolling mills & improvements.

9. a) Question in natural language.

Improved blast furnace stove design.

b) UDC Translation.

669.162.23) hot blast stoves
669.162.238.5)

c) Keyword Translation.

Blast furnaces & stoves.

10. a) Question in natural language.

Refractory coating of oxygen lances.

b) UDC Translation.

666.76) Refractories & 666.76) Refractories
666.76) 666.767)

669.046.564.5) Oxygen lances & 669.183.212.6) OH
669.046.564.5) 669.183.212.65) roof
lances

c) Keyword Translation.

Lances & refractories
Oxygen lances & refractories
Lances & magnesite
Lances & basic refractories
Lances & chromite refractory
Lances & siliceous
Oxygen lances & magnesite
Oxygen lances & basic refractories
Oxygen lances & chromite refractory
Oxygen lances & siliceous

It will be noted that in general there is not much difference in the question space occupied by the UDC and Keywords versions. The questions were input by means of punched cards.

Numerically the results were

	UDC	Keywords
1	33	19
2	1	-
3	3	-
4	1	-
5	1	1
6	-	4
7	-	2
8	-	-
9	3	1
10	3	1

The actual abstracts retrieved in respect of each of the 9 questions which provided output are given in Appendix 1.

It could be proposed that these results provide evidence that the UDC is more effective in computer based information retrieval systems than keywords. However, there are three major factors to be considered in this instance. These are

1. The BISRA Theasaurus used was in the early stages of development.
2. The staff involved in operating the system, while experienced information workers, had no great experience of keyword indexing.
3. The version of the UDC used was ISI Special Report No.84, UDC; Special Subject Edition for Metallurgy, which, without doubt, is one of the better sections of the UDC.

So perhaps it could be argued that we are not really comparing like with like. However, it is quite clear from these preliminary results that the UDC can be used effectively in an interrogation system and it is our intention to carry out further work in this area. While we have stopped adding to the data base it is available to us for further experimental work. It is unlikely that we will undertake any further interrogation until we transfer the data base to disc (from tape).

Acknowledgements.

I wish to make it clear that this project involved co-operation by the following.

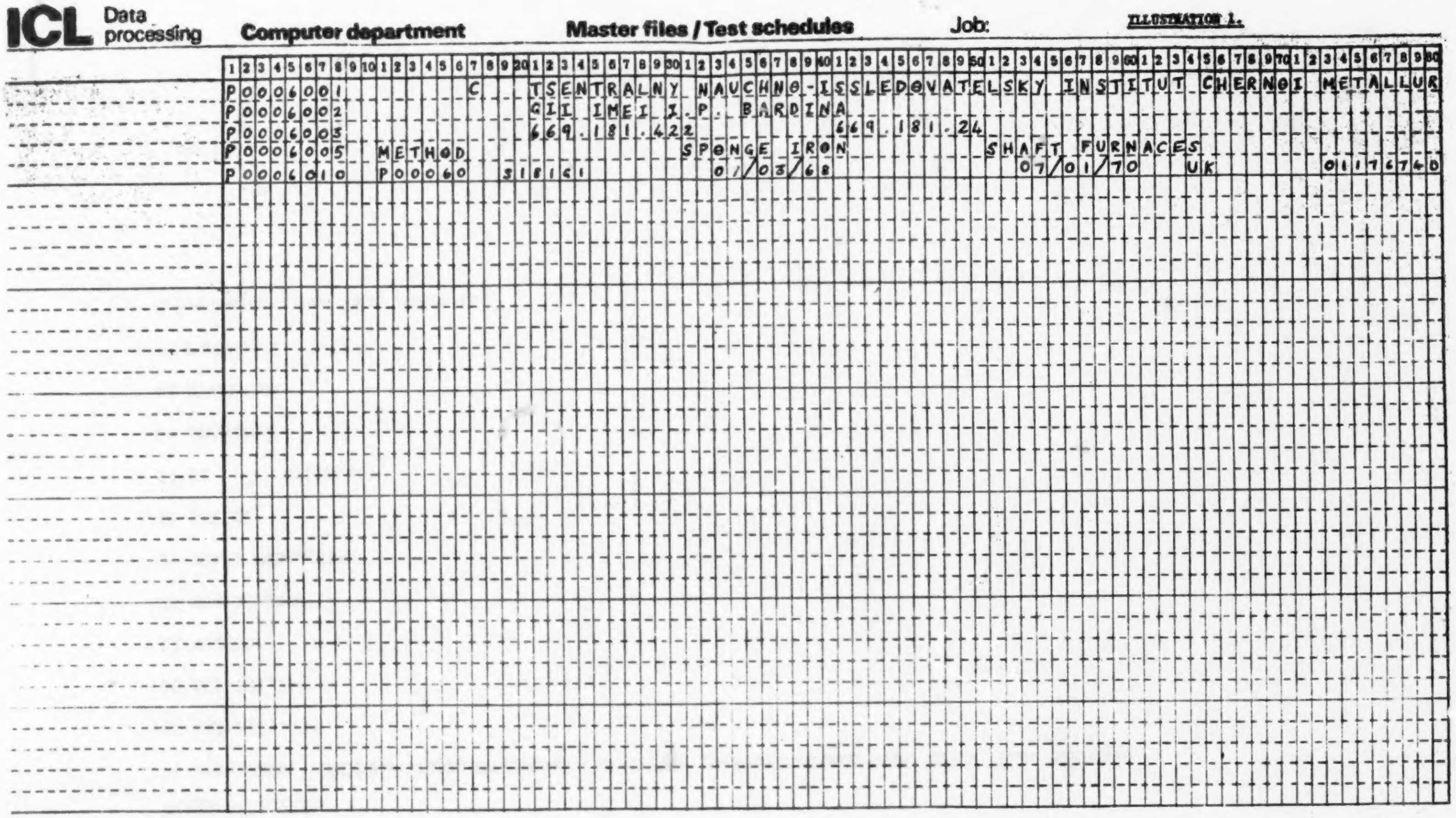
K.W. Atkin	BISRA
M.J. Hoolahan	BSC Headquarters
M.N. Patten	Strip Mills Division
D.A. Pleasance	BISRA
R. Sewell	Special Steels Division
J.C. Sykes	formerly of BISRA now BSC Headquarters
K.T.C. Tuck	Tubes Division

12

The original free text search programme was created by E. Pace, M.Sc , when he was with the British Steel Corporation, while J.J. Carmichael and W. Baxter of the BSC General Steels Division Library and Information Services were responsible for the week to week creation and operation of the system. Gratitude is also due to the staff of the Department of Computer Operations, Scottish Shelton and East Moors Group, BSC for their assistance during the whole project.

Also I wish to express my gratitude to the British Steel Corporation for the generous support given to this work which made possible my participation in this Seminar.

ILLUSTRATIONS TO THE PAPER OF MR.R.HINDSON



ICL Data processing **Computer department** **Master files / Test schedules** Job: ILLUSTRATION 1.

```
P0006001        C   TSENTRALNY NAUCHNO-ISSLEDOVATELSKY INSTITUT CHERNOI METALLUR
P0006002            GII IMEI I.P. BARDINA
P0006003            669.181.422         669.181.24
P0006005  METHOD              SPONGE IRON         SHAFT FURNACES
P0006010  P00060  318161        01/03/68            07/01/70   UK       0117674D
```

<u>Illustration 2.</u>

```
06 TSENTRALNY  NAUCHNΘ- ISSLEDΘVATELSKY  INSTITUT  CHERNΘI  METALLURGII  IMEI  I.P. BARDINA↑
21 669.181.422↑
22 669.181.24↑
27 METHΘD↑
28 SPΘNGE IRΘN↑
29 SHAFT FURNACES↑
52 SIBIGI↑
57 UK      01176740↑
64 07/03/68↑
66 07/01/70↑
99 P00060↑
```

↑represents a new line character.

Illustration 3.

```
01 A METHØD ØF ØBTAINING A SPØNGY METAL, FØR EXAMPLE SPØNGE IRØN,+
02 INCLUDING THE STEPS OF MØVING METAL ØXIDES AND A SØLID REDUCING+
03 AGENT IN IMMISCIBLE FLØW STREAMS IN THE SHAFT ØF AN EXTERNALLY+
04 HEATED SHAFT FURNACE IN SUCH A MANNER THAT THE FLØW STREAM ØF+
05 THE METAL ØXIDES IS INSIDE THE FLØW STREAM ØF THE SØLID REDUCING+
06 AGENT SØ THAT THE ØXIDES ARE REDUCED BY THE SØLID REDUCING AGENT+
99 P00060+
```

ILLUSTRATION 4

Tag Numbers for Paper Tape Input for Elements of Bibliographic Description of Patents

01	individual authors (19 characters)
02	
03	
04	
05	
06	corporate author (99 chs.)
07	et al
21	UDC numbers (19 chs.)
22	
23	
24	
25	
27	keywords (19 chs.)
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
52	source indicator
57	patent number
64	earliest patent date
65	middle patent date
66	latest patent date
99	closure of record.

Note:- missing numbers are not required by patent system.

```
IRON.

UK       01173991
   EMHART CORP.
   666,763.3             666,763.31          621,744
   536,421.3
        REFRACTORIES        MIXING              ZIRCONIA
        ALUMINIUM OXIDE     MULLITE             SLIP CASTING
        SINTERING
            P00026  S1               04/03/66  24/02/67  01/01/70
 A REFRACTORY BODY IS OBTAINED BY MIXING IN WEIGHT PER CENT, ZIRCON
 3,6, MOULDING, E.G. BY SLIP CASTING AND THEN SINTERING AT 1400 -
 1500C.

UK       01173994
   TAYLOR AND CHALLEN LTD.
   621.9.062             62-82
        PRESSES             DIES                HYDRAULIC SYSTEMS
        SHOCK ABSORBERS
            P00002  B1               07/03/67            01/01/70
 A PRESS HAVING A BOTTOM DIE OR PLATEN AND PROVIDED WITH A DIE
 CUSHION COMPRISING A PISTON CONNECTED TO THE BOTTOM DIE OR PLATEN
 AND SLIDABLE IN A CYLINDER, UNDER THE EFFECT OF PRESSURE ON THE
 DIE OR PLATE, TO DISPLACE A HYDRAULIC FLUID FROM THAT CYLINDER
 THROUGH A FLOW-THROTTLING PASSAGE INTO A RESERVOIR COMPRISING A
 CLOSED SPACE CONTAINING A HYDRAULIC FLUID ABOVE THE SURFACE LEVEL
 OF WHICH THERE IS TRAPPED A BODY OF GAS IN DIRECT CONTACT WITH
 THE HYDRAULIC FLUID.

UK       01173998
   COMPAGNIE FRANCAISE THOMSON HOUSTON-HOTCHKISS BRANDT
   531.71                53,082.5            53,082.7
        DIMENSIONS          MEASUREMENT         OPTICAL MEASUREMENT
        ELECTRICAL MEASURES IMPROVEMENT         METHOD
            P00012  B1               210766FR  21/07/67  01/01/70
 UPTO-ELECTRICAL METHOD OF MEASURING AND REMOTE-CHECKING A
 DIMENSION OF A MEMBER CHARACTERISED IN THAT IT USES TWO FINE, NON
 (OR ONLY SLIGHTLY) DIVERGENT, PARALLEL LIGHT BEAMS REFLECTED AT
 TWO POINTS ON THE MEMBER WHICH DEFINE THE SAID DIMENSIONS, THE
 TWO BEAMS BEING MODULATED SINUSOIDALLY IN INTENSITY AND IN PHASE
 CONCORDANCE IN TWO OPERATIONS ALONG THE OPTICAL PATH, A FIRST TIME
 ON THE INCIDENT PATH AND A SECOND TIME ON THE RELFLECTED PATH, THE
 LENGTHS OF THE OPTICAL PATHS COVERED BY BOTH BEAMS BETWEEN THE
 TWO MODULATION OPERATIONS BEING RENDERED RECURRENTLY AND
 CONTINUOUSLY VARIABLE ACCORDING TO A SIMPLE LAW, THE AMPLITUDE OF
 THIS VARIATION BEING EQUAL TO A WAVE-LENGTH OF THE MODULATION
 SIGNAL, THE SUCCESSIVE MODULATION CAUSING THE APPEARANCE OF A
 RECURRENT VIDEO COMPONENT, INDEPENDENT OF THE MODULATION FREQUENCY
 AND A FUNCTION OF THE LENGTH OF THE OPTICAL PATH.

UK       01176012
   DOW CHEMICAL CO.
   620.197.2             66.074.3
        CORROSION           PREVENTION          METALS
        GAS CLEANING        GAS FLOW            GASES
        CONTAMINATION       ACIDS               PLANTS
            P00417  T1               190166US  17/01/67  01/01/70
 CORROSION PREVENTION OF METALS IN CONTACT WITH ALKANOLAMINE SOLNS.
 USED TO SCRUB GAS STREAMS FREE FROM ACID GAS CONTAMINANTS OF A
 METAL SCRUBBING PLANT.  TO THE ALKANOLAMINE SOLN. IS ADDED UP
```

```
        METHOD                 CONTROL SYSTEMS      REFINING
        PIG IRON               BASIC CONVERTERS     INTENSITY
        FLAMES                 RADIATION            MEASUREMENT
        DECARBURIZING          RATE                 CURVES
          P00350  S1B1G182       180467LU  18/04/68  04/03/70    UK      01185281
CENTRE NATIONAL DE RECHERCHES METALLURGIQUES
    669.184.236           669.184.244.66
        PIG IRON               REFINING             OXYGEN STEELMAKING
        EXHAUST GASES          COMBUSTION
          PU0063  S2S10161       230866LU  16/06/67  07/01/70    UK      01176809
CENTRO SPERIMENTALE METALLURGICO S.P.A.
    669.184.2             669.13-154          669.054.82
        REFINING               IRON ORES            PROCESSES
        PRODUCTION             STEELS               LOW ALLOY STEELS
        SULPHUR                PHOSPHOROUS          OXYGEN STEELMAKING
        CAST IRON              MELTS                CONVERTERS
        SLAGS                  MIXTURES
          PC0085  B1T161         200367IT  04/03/68  28/05/70    UK      01193391
CENTRO SPERIMENTALE METALLURGICO S.P.A.
    669.184.244.66
        PROCESSES              CONTINUOUS PROCESS  CONVERSION
        CAST IRON              STEELS               REACTIONS (CHEMICAL)
        REFINING               FLOW                 DISPERSING
        PHASES                 SLAGS                EFFERVESCENCE
        EMULSIONS              STIRRING             OXYGEN
        JETS
          P00683  S1B161         180367IT  06/03/68  15/04/70    UK      01187942
CESKOSLOVENSKA AKADEMIE VED.
    543.544               547.466             547.583.5
        CHROMATOGRAPHY         ANALYZING            MIXTURES
        AMINO ACIDS            AUTOMATION
          P00476  T1             070166CZ  05/01/67  14/01/70    UK      01177781
CHARBONNAGES DE FRANCE AND HOLUILLESES.
    622.781
        HIGH TEMPERATURE       HEAT TREATMENT       GRANULAR MATERIALS
        SOLIDS                 FLUIDIZING           BEDS
        SINTERING              PELLETS              AGGLOMERATION
        REACTORS               GRATE                CORES
          PU0701  T1             290766FR  29/07/67  18/03/70    UK      01184738
CHASE BRASS AND COPPER CO. INC.
    621.771.237           621.771.06-589
        EQUIPMENT              SHEET                GUIDES
        TENSION ROLLING
          P00113  B1             180266US  16/02/67  14/01/70    UK      01178021
CHELTON (FORMING) LTD.
    621.743.073
        IMPROVEMENT            PRODUCTION           CORES
        BOXES                  FOUNDRIES            COATING
        METAL SPRAYING
          PU0374  B1             07/03/67            11/03/70    UK      011840[illegible]
CHELYABINSKY METALLURGICHESKY ZAVOD
    621.365.2             669.046.585
        ELECTRIC ARCS          FURNACES             SLAGS
        MELTING
          P00998  T1B161         04/10/68            03/04/70    UK      01193527
CHELYABINSKYMETALLURGICHESKY ZAVOD
    669.187.28
        IMPROVEMENT            ELECTRIC ARCS        FURNACES
        SLAGS                  METALS               COOLING
        LININGS                WALLS                MELTS
```

```
      PLASTICS              PIPES                  JOINTS
      SEALING               RINGS                  GROOVE
        P00594  T1              19/02/68  18/02/69  11/03/70    UK      01183881
621.643.4           678.058
    SMITH A O CORP.
      HEATING EQUIPMENT     CURING                 PLASTICS.
      PIPES                 JOINTS                 ADHESIVES
      TUBULAR               ALUMINIUM              HEATED
      INSULATION            RINGS                  APPARATUS
      GASES                 JETS                   TUBES
        P00747  T1              270566US  24/05/67  25/03/70    UK      01185864
621.643.4           678.073
    BTR INDUSTRIES LTD.
      PIPES                 JOINTS                 FLANGES
      LININGS               THERMAL PROPERTIES  PLASTICS
        P00695  T1              11/05/66  05/05/67  04/02/70    UK      01180392
621.643.4           697.32
    ENERJY S.A.
      PIPES                 JOINTS                 PUMPS
      BOILERS               VALVES                 HOUSINGS
      HEATING               SYSTEMS                RINGS
      ELASTIC DEFORMATION HOLDERS                  BOLTED JOINTS
      SEALING
        P00742  T1              06/08/68            25/03/70    UK      01185766
621.7               677.46
    ROTARY PROFILE ANSTALT.
      PROCESSING            PLASTIC DEFORMATION INCLUSIONS
      FIBRES                HEATING                ISOLATION
      TORQUE
        P00613  S101            0505662A  21/04/67  02/04/70    UK      01186201
621.7-72            621.892
    MOBIL OIL CORP.
      METALWORKING          LUBRICANTS             EMULSIONS
      LUBRICATING OILS
        P00018  B1              270566US  26/05/66  01/01/70    UK      01176076
621.7.016.3         621.892
    BADISCHE ANILIN AND SODA FAB.
      SOLUTIONS             RESINS                 LUBRICANTS
      THERMAL PROPERTIES  PLASTICS                 METALS
      CORROSION             INHIBITORS             DEEP DRAWING
      TUBEMAKING            COLD ROLLING
         00564  T1              0206668F  01/04/67  04/03/70    UK      01183065
621.7.022.2         620.197.6
    WOOD W A            GRAY D
      PROTECTIVE COATINGS PRIMERS (COATINGS)
        P00522  T1              06/01/66  06/01/67  18/02/70    UK      01181201
621.7.023.2         621.791.55             669.788                660.14-613
    TODD J P
      FUSION WELDING        FABRICATION            PLATE
      STRUCTURAL STEELS   SHOT BLASTING            SPRAYING
      RESINS                PRIMERS(COATINGS)   EPOXY RESINS
      ALUMINIUM             POWDERS                DRYING
      ORGANIC COMPOUNDS
        P00251  S1              12/06/67            11/02/70    UK      01180964
621.7.026
    DOW CHEMICAL CO.
      METAL CLEANING        ADHESION               HEATING
      TEMPERATURE           CONTACTING             LIQUIDS
      INDUCTION HEATING   COILS                    VACUUM
      SUBMERGING
```

APPENDIX TO THE PAPER OF MR.R. HINDSON

Question 1 :

NEW PROCESSES FOR PRODUCTION OF STEEL (UDC)

```
UK      01176740
   TSENTRALNY NAUCHNO-ISSLEDOVATELSKY INSTITUT CHERNOI METALLUR
   GII IMEI I.P. BARDINA
   669.181.422          669.181.24
       METHOD                  SPONGE IRON           SHAFT FURNACES
           P00060  818161          07/03/68                07/01/70
  A METHOD OF OBTAINING A SPONGY METAL, FOR EXAMPLE SPONGE IRON,
  INCLUDING THE STEPS OF MOVING METAL OXIDES AND A SOLID REDUCING
  AGENT IN IMMISCIBLE FLOW STREAMS IN THE SHAFT OF AN EXTERNALLY
  HEATED SHAFT FURNACE IN SUCH A MANNER THAT THE FLOW STREAM OF
  THE METAL OXIDES IS INSIDE THE FLOW STREAM OF THE SOLID REDUCING
  AGENT SO THAT THE OXIDES ARE REDUCED BY THE SOLID REDUCING AGENT.

UK      01179958
   WELLMAN ENGINEERING CORP.
   669.18-932
       SPRAY REFINING        IRON                   FLOW MEASUREMENT
       LADLES                TUNDISH                CARBON
       SILICON               ANALYZING
           P00188  8161          20/06/67               06/02/70
  A METHOD OF MEASUREMENT OF IRON IN SPRAY STEELMAKING, COMPRISING
  MEASURING THE WEIGHT OF IRON IN AN IRON-RECEIVING LADLE, USING THE
  SAID WEIGHT MEASUREMENT IN RELATION TO A PREDETERMINED TIME
  INTERVAL TO PROVIDE AN INDICATION OF THE FLOW-RATE FROM THE SAID
  LADLE, MEASURING AND INDICATING THE HEIGHT OF METAL IN A TUNDISH
  OR OTHER RECEPTACLE INTERPOSED BETWEEN THE IRON-RECEIVING LADLE
  AND THE SPRAY STEELMAKING UNIT, MEASURING THE TEMPERATURE OF THE
  METAL IN THE SAID TUNDISH OR OTHER RECEPTACLE, APPLYING THE SAID
  HEIGHT AND TEMPERATURE MEASUREMENTS TOGETHER WITH A SIGNAL BASED
  UPON THE KNOWN CARBON SILICON ANALYSIS OF THE IRON, TO AN
  INSTRUMENT WHICH PROVIDES AN OUTPUT SIGNAL PROPORTIONAL TO THE
  FLOW-RATE FROM THE SAID TUNDISH OR OTHER RECEPTACLE AND APPLYING
  SAID OUTPUT SIGNAL TO AN INDICATOR WHICH PROVIDES AN INDICATION OF
  THE SAID FLOW-RATE FROM THE TUNDISH OR OTHER RECEPTACLE IN FLOW
UNITS

UK      01184278
   GEBR BOHLER AND CO. A.G.
   669.054.7
       INSTALLING            ELECTROSLAG PROCESS METALS
       CONSUMABLE            ELECTRODES          MOULDS
       LIQUID METALS         ELECTRIC CURRENT    FLOW
           P00404  8161          061167AU  06/11/68  11/03/70
  AN INSTALLATION FOR THE ELECTRO SLAG REMELTING OF METALS, HAVING
  A MAIN ELECTRODE CONSISTING OF CONSUMABLE METAL, A MOULD AND AN
  AUXILIARY ELECTRODE SERVING FOR MAINTAINING THE HEAT IN THE
  MOLTEN METAL FORMED, AND THROUGH WHICH CURRENT FLOWS DURING THE
  PROCESS OF THE EXCHANGING OF THE MAIN ELECTRODE, CHARACTERISED IN
  THAT THE AUXILIARY, AND/OR THE MOULD, AND THE MAIN ELECTRODE,
  ARE ARRANGED SO AS TO BE MOVABLE ON TRANSPORT DEVICES WHICH ARE
  SEPARATE FROM EACH OTHER.

UK      01175982
   ASSOCIATED ELECTRICAL INDUSTRIES LTD.
   669.14-462          669.14-412          669.054.7
       PRODUCTION            TUBULAR               METALS
       INGOTS                ELECTROSLAG PROCESS COOLING
       LIQUIDS               FLOW                  ELECTRODES
       MOULDS                ALTERNATING CURRENT
```

```
                                                                        10/06/
                   NEW PROCESSES FOR PRODUCTION OF STEEL (UDC)             2
           P00416  T1                 19/04/67  19/04/68  01/01/70
  TUBULAR METAL INGOTS ARE PRODUCED BY AN ELECTROSLAG PROCESS IN AN
  ANNULAR MOULD.  COOLING LIQUID FLOWS DOWN THE OUTER SURFACE OF THF
  OUTER WALL AND THE INNER SURFACE OF THE INNER WALL INTO AN ANNULAR
  TROUGH.  ELECTRODES AT SPACED LOCATIONS AROUND THE MOULD ARE
  SUPPORTED IN GROUPS OF THREE TO WHICH A THREE-PHASE A.C. CURRENT
  IS SUPPLIED .  A COOLING LIQUID DRAIN FROM THE CENTRAL REGION IS
  SURROUNDED BY A BARRIER TO FORM AN ANNULAR CHANNEL INTO WHICH
  COOLING LIQUID MAY BE INTRODUCED SEPARATELY.  A BARRIER WITHIN THE
  TROUGH FORMS AN ANNULAR CHANNEL WITH THE MOULP INTO WHICH THE
  COOLING LIQUID MAY BE INTRODUCED SEPARATELY.

UK      01184877
    STANDARD MESSO DUISBURG GESELLSCHAFT
    669.18-932
        DROPS                OXYGEN              GASES
        PIG IRON             STEELS              PRODUCTION
        APPARATUS            CONTINUOUS PROCESS  EVACUATION
           P00442  B1                 190667GE  18/06/68  18/03/70
  A PROCESS FOR THE CONTINUOUS PRODUCTION OF STEEL FROM PIG IRON,
  IN WHICH THE IRON IS INTRODUCED IN A STREAM INTO AN
  EVACUATED GAS REACTION CHAMBER, DIVIDED INTO DROPLETS AND THEN
  TREATED WITH OXYGEN.

UK      01188867
    EMPRESA AUXILIAR DE LA INDUSTRIA S.A.
    622.366.12           669.141.1
        PROCESSES            UTILIZATION         IRON
        PYRITE               COMBUSTION PRODUCTS LEACHING
        PELLETS              MIXTURES            NON-FERROUS METALS
           P00729  B161               300967ES  12/08/68  22/04/70
  A PROCESS FOR ULTIZATION OF THE IRON CONTAINED IN PYRITES CINDERS,
  COMPRISING LEACHING PYRITES CINDERS WITH LEACHING LIQUOR TO
  OBTAIN PURPLE ORE AND AN IRON-CONTAINING SOLUTION, S[illegible]RATING NON -
  FERROUS METALS FROM THE IRON CONTAINED IN AT LEAST PART OF THE
  [illegible]-CONTAINING SOLUTION, INCORPORATING WITH THE PURPLE ORE AT
  LEAST PART OF THE IRON CONTAINED IN THE IRON-CONTAINING SOLUTION,
  AND FORMING PELLETS FROM THE MIXTURE THUS OBTAINED.

UK      01190075
    ASSOCIATED ELECTRICAL INDUSTRIES LTD.
    669.054.7            669.14.412
        IMPROVEMENT          ELECTROSLAG PROCESS PLANTS
        INGOTS               PRODUCTION          ELECTRODES
        MELTING              ELECTRIC CURRENT    VARIATIONS
           P00775  B16131             07/04/67            29/04/70
  ELECTROSLAG PLANT FOR INGOT PRODUCTION FROM A PLURALITY OF PROCESS
  ELECTRODES CONNECTIBLE TO A COMMON ELECTRICAL SUPPLY FOR
  CONTEMPORANEOUS MELTING, THE PLANT BEING ARRANGED FOR CONTROL OF
  THE PROGRESSIVE FEEDING OF THE ELECTRODES DURING THE PROCESS BY
  A SINGLE CONTROLLER RESPONSIVE TO ELECTRODE CURRENT VARIATIONS.

UK      01193457
    CONZINC RIOTINTO OF AUSTRALIA LTD.
    669.046.512-932      669.046.55          669.141.1
        CONTINUOUS PROCESS   SMELTING            IRON ORES
        AGGLOMERATES         REFINING            BATHS
        FURNACES             SLAGS               LIQUID METALS
        PIG IRON             IRON-BASE ALLOYS
           P00992  T161B1             281166AA  23/11/67  03/06/70
```

```
                 NEW PROCESSES FOR PRODUCTION OF STEEL (UDC)
 A CONTINUOUS SMELTING PROCESS IN WHICH OXIDIC IRON ORE IN
 AGGLOMERATED OR PARTICULATE FORM IS FED INTO THE SMELTING
 ZONE OF A CONTINUOUS BATH SMELTING-REFINING FURNACE HAVING A
 SMELTING ZONE, A REFINING ZONE CONNECTED TO THE SMELTING ZONE
 AND A SLAG SEPARATION ZONE MOLTEN MATERIAL IS REFINED IN THE
 REFINING ZONE AND SLAG IS SEPARATED FROM METAL IN THE SLAG
 SEPARATION ZONE, CHARACTERIZED IN THAT A CARBON REDUCTANT
 PART AT LEAST OF WHICH IS IN THE FORM OF MOLTEN PIG IRON,
 SEMI-STEEL OR OTHER CARBON-CONTAINING FERROUS ALLOY IS
 CONTINUOUSLY FED INTO THE SMELTING ZONE DIRECTLY OR VIA THE
 SLAG SEPARATION ZONE.

UK      01194948
   GEBR, BOHLER AND CO. A.G.
   669,054,7            621.3.036.3
        APPARATUS              JOINING              LINES
        ELECTRICITY            ELECTRODES           PLATE
        INSTALLING             ELECTROSLAG PROCESS METALS
           P01076  01               205167AU  20/11/68  17/06/70
 A DEVICE FOR CONNECTING ELECTRIC LINES TO ELECTRODES, BASE PLATES,
 OR SIMILAR CONDUCTING ELEMENTS, IN INSTALLATIONS FOR THE ELECTRO-
 SLAG REMELTING OF METALS, IN WHICH TWO PIVOT LEVERS ARE PROVIDED,
 MOUNTED OPPOSITE EACH OTHER AND ROTATABLE IN THE PLACE DEFINED BY
 THEM, CARRYING IN EACH CASE AT LEAST ONE CLAMP JAW, BETWEEN WHICH
 PIVOT LEVERS THERE IS CONNECTED A PISTON CO-OPERATING WITH A
 PNEUMATIC OR HYDRAULIC CYLINDER, BY MEANS OF WHICH THE CLAMP
 JAWS CAN BE PRESSED ON THE CONDUCTING ELEMENT WHICH IS TO BE
 ATTACHED AND CAN BE RELEASED FROM THE CONDUCTING ELEMENTS WHICH
 HAS BEEN ATTACHED, THE PISTON BEING CONNECTED TO ONE OF THE PIVOT
 LEVERS, AND THE CYLINDER BEING CONNECTED TO THE OTHER PIVOT LEVER.

UK      01196186
   J.C.I. METAL HOLDINGS PROPRIETARY LTD.
   669,181,272          622,341,1            622,788
        IRON ORES              PELLETS              COKE
        GRANULATION            SHAFT FURNACES       REDUCTION (GENERAL)
        DISCHARGE              SCREENING (SIZING)   MOLTEN METALS
        STEEL                  DECARBURIZING        DRYING
        GASES                  FLUXES
           P01113  S1T16181         06/02/68  06/02/69  24/06/70
 PELLETS OF FINELY-DIVIDED IRON ORE AND CARBONACEOUS REDUCING AGENT
 HAVE MIXED WITH THEM GRANULAR CARBONACEOUS MATERIAL OF SMALLER
 PARTICLE SIZE THAN THE PELLETS AND ARE THEN HEATED IN A SHAFT
 KILN SO  THAT THE PELLETS ARE AT LEAST 90% REDUCED; AT THE BOTTOM
 OF THE SHAFT THE MIXTURE PASSES OVER A DISCHARGE EDGE TO A SCREEN
 OUT OF CONTACT WITH THE AIR, WHERE THE CARBONACEOUS MATERIAL IS
 SEPARATED FROM THE REDUCED PELLETS, THE LATTER PASSING TO A FURNACE
 CONTAINING A BATH OF STEEL WHERE THEY ARE MELTED WHILE THE FORMER
 IS COOLED TO BELOW COMBUSTION TEMPERATURE AND RECYCLED. BY
 REMOVING CARBON FROM THE MIXTURE IN THIS WAY STEEL MAY BE DIRECTLY
 PRODUCED IN THE MELTING FURNACE WITHOUT THE NECESSITY OF A
 CONVERSION OPERATION. AS ILLUSTRATED, MOIST PELLETS FROM HOPPER
 ARE DRIED AND PRE-HEATED BY HOT GASES WHILE TRAVELLING ON CONVEYOR
 TO A HOPPER, WHERE THEY ARE MIXED WITH GRANULAR CARBONACEOUS
MATERIALS.

UK      01196186
   J.C.I. METAL HOLDINGS PROPRIETARY LTD.
   669,181,272          622,341,1            622,788
        IRON ORES              PELLETS              COKE
```

```
                                                                          10/06/
                  NEW PROCESSES FOR PRODUCTION OF STEEL (UDC)               4
      GRANULATION           SHAFT FURNACES       REDUCTION (GENERAL)
      DISCHARGE             SCREENING (SIZING)   MOLTEN METALS
      STEEL                 DECARBURIZING        DRYING
      GASES                 FLUXES
          P01113  S1T1G1S1        06/02/68  04/02/69  24/06/70

UK      01198238
   HUTTENWERK OBERHAUSEN A.G.
   66.044.14            669.181.24
      SHAFT FURNACES        REDUCTION (CHEMICAL PARTICLES
      GRANULAR MATERIALS    IRON ORES            REDUCING AGENTS
      GASES                 CHARGING             HOPPERS
      WALLS                 DISCHARGE            PIPES
      VALVES                PLATE
          P01191  S1G1S1          10/07/68  27/06/69  08/07/70
  A RECTANGULAR SHAFT FURNACE FOR THE REDUCTION OF PARTICULATE IRON
  ORE WITH A REDUCING GAS COMPRISES A CHARGE HOPPER, A RIDGED GAS
  DISTRIBUTER DISPOSED PARALLEL TO THE LONGER WALLS OF THE SHAFT
  ABOVE A HOPPER DISCHARGE BASE WITH WALLS PARALLEL TO THE
  DISTRIBUTER TAPERING TO A SLOT CORRESPONDING IN WIDTH TO SAID
  DISTRIBUTER, AND A DISCHARGE PLATE OVER WHICH RECIPROCATES ONE
  ORE MORE PUSH/PULL BEAMS. THE BREADTH OF THE GAS DISTRIBUTER IS
  1/6 - 1/3, PREFERABLY 1/4, THAT OF THE WALL AT RIGHT ANGLES TO IT,
  AND THE RIDGE TAPER ANGLE LIES BETWEEN 40 AND 50 DEGREES. PIPES
  SUPPLY GAS TO BOTH ENDS OF DISTRIBUTER. THE WALLS ARE INCLINED
  AT 20 - 40 DEGREES TO THE VERTICAL.

UK      01198290
   SOCIETE METALLURGIE D'IMPHY
   621.74.047           669.18-932
      JOULE EFFECT          HEATING              FEEDING
      FLOW                  PROCESSES            CONTINUOUS PROCESS
      CONTINUOUS CASTING    REFINING             METALS
      LAYERS                ELECTRICITY          CONDUCTIVITY
      SLAGS                 LIQUID METALS        COOLING
      CASTING               MOULDS
          P01214  S1S1            060167FR  05/12/67  08/07/70
  A PROCESS FOR CONTINUOUSLY CASTING AND REFINING A METAL,
  COMPRISING THE STEPS OF FORMING A LAYER OF AN ELECTICALLY
  CONDUCTIVE SLAG ABOVE THE MOLTEN METAL LEVEL IN A COOLED,
  BOTTOMLESS, CASTING MOULD, DIRECTING A STREAM OF MOLTEN METAL
  THROUGH THE SLAG LAYER TO FEED THE MOLTEN METAL WITHIN THE MOULD
  HEATING BY THE JOULE EFFECT THE SLAG TO MAINTAIN IT IN LIQUID
  FORM, AND ADDING FRESH SLAG TO, AND REMOVING SLAG FROM, THE
  LAYER TO MAINTAIN THE EFFECTIVENESS OF THE SLAG IN THE LAYER.

UK      01199703
   METALLGESELLSCHAFT A.G.
   669.181.25           669.181.4
      PROCESSES             PRODUCTION           SPONGE IRON
      SULPHUR               IRON ORES            DIRECT REDUCTION
      KILNS                 SOLID STATE          CARBON ALLOYS
      REDUCING AGENTS
          P01279  S1T1O1          090268GE  27/01/69  22/07/70
  A PROCESS FOR PRODUCING SPONGE IRON HAVING A LOW SULPHUR CONTENT,
  WHEREIN A MATERIAL CONTAINING IRON OXIDE IS DIRECTLY REDUCED IN
  A ROTARY KILN WITH THE AID OF A SOLID CARBONACEOUS REDUCING AGENT
  WHICH IS SUPPLIED INTO THE ROTARY KILN AT THE DISCHARGE END
  THEREOF, AND WITH THE AID OF A SULPHUR-COMBINING SUBSTANCE, AND
  WHEREIN AT LEAST PART OF THE SULPHUR-COMBINING SUBSTANCE WHICH IS
```

```
REQUIRED IS CHARGED DIRECTLY INTO THE REDUCING ZONE OF THE
ROTARY KILN AND THE TOTAL WEIGHT OF THE CHARGED SULPHUR-
COMBINING SUBSTANCE IS AT MOST 4% OF THE WEIGHT OF THE CHARGED
ORE.

UK      01200278
   BOHLER AND CO. A.G. GEB
   669.034.7            669.14-412
        IMPROVEMENT           INSTALLING          MANUFACTURING
        INGOTS                VARIATIONS          DIMENSIONS
        ELECTROSLAG PROCESS   STEELS              ELECTRODES
        HOLDERS               MOULDS              CONSUMABLE
           P01298  B1B1             141267AU  13/12/68  29/07/70
  AN INSTALLATION FOR THE MANUFACTURE OF INGOTS OF DIFFERENT SIZES
  IN THE ELECTRO-SLAG REMELTING OF METALS, IN PARTICULAR OF STEELS,
  CHARACTERIZED IN THAT THERE ARE PROVIDED TWO ELECTRODE HOLDERS
  DISPLACEABLE OR PIVOTABLE IN THE HORIZONTAL DIRECTION AND TWO
  MOULDS, WHEREIN THE ELECTRODE HOLDERS IN SEQUENCE ALTERNATELY
  SUPPLY CONSUBABLE ELECTRODES TO ONE OF THE TWO MOULDS, AND ONE OF
  THESE ELECTRODE HOLDERS INSERTS CONSUMABLE ELECTRODES INTO THE
  OTHER OF THE TWO MOULDS.

UK      01200409
   INSTITUT DE RECHERCHES DE LA SIDERURGIE FRANCAISE
   66.073               669.18-932
        PROCESS               APPARATUS           CONTINUOUS PROCESS
        METALS                REFINING            CONTAINERS
        SLAGS                 OUTPUT              FEEDING
        DUCT                  EXTRACTION          GASES
        CONTROL SYSTEMS       LOSSES
           P01301  B1B1G1T1         180967FR  06/09/68  29/07/70
  IN THE OPERATION OF A CONTINUOUS METAL REFINING UNIT COMPRI[illegible]
  A SERIES OF RECEPTACLES OF WHICH THE FIRST, CALLED THE REAC[illegible]
  CONTAINER, COMMUNICATES WITH THE LAST RECEPTACLE EQUIPPED [illegible]
  A SLAG OUTFEED OPENING COMMUNICATING WITH THE ATMOSPHERE, [illegible]
  REACTION CONTAINER AND THE LAST RECEPTACLE BEING LINKED TO [illegible]
  FOR EXTRACTING THE GASES PRODUCED DURING THE REFINING REACTIO[illegible]
  A PROCESS FOR ENSURING THIS EXTRACTION WITHOUT GAS LOSS THROUGH
  THE SAID SLAG REMOVAL OPENING, IN WHICH PROCESS THERE IS CONTROL
  OF THE AMOUNT Q OF GASES EXTRACTED INTO THE EXTRACTION TUBE
  LINKED TO THE REACTION CONTAINER, TO A VALVE LOWER THAN THE
  AMOUNT Q OF GASES PRODUCED IN NORMAL OPERATION DURING REFINING
  REACTIONS AND THERE IS CORRELATED CONTROL OF THE AMOUNT OF GASES
  EXTRACTED INTO THE SUCTION DUCT LINKED TO THE LAST RECEPTACLE.

UK      01200880
   SWISS ALUMINIUM LTD.
   669.034.7            66.074
        METHOD                APPARATUS           REMOVAL
        EXHAUST GASES         ELECTROCHEMISTRY    FURNACES
        PUMPS
           P01340  B1               300768SW  15/07/69  05/08/70
  A METHOD OF REMOVING THE EXHAUST GASES FROM AN OPEN ELECTROLYTIC
  FURNACE IN A FURNACE HOUSE IN WHICH AN AIR CURTAIN DIRECTED
  TOWARDS A SUCTION APPARATUS IS SET UP AROUND THE FURNACE TO
  PRECENT THE ESCAPE OF THE EXHAUST GASES INTO THE ATMOSPHERE
  INSIDE THE FURNACE HOUSE.

UK      01202192
   BRITISH IRON AND STEEL RESEARCH ASSOCIATION
```

FILMED FROM BEST AVAILABLE COPY

```
                                                                   10/06,
                  NEW PROCESSES FOR PRODUCTION OF STEEL (UDC)          6
  669.054.7            621.3.033.2           621.744.3
      IMPROVEMENT               METHOD                 ELECTROSLAG PROCESS
      MOTION                    MOULDS                 ELECTRODES
      MELTING                   APPARATUS
          P01366  B16181           13/09/66              12/08/70
 A METHOD OF ELECTRO-SLAG REFINING IN WHICH NON-RECIPROCATING
 RELATIVE MOTION IS EFFECTED BETWEEN A MOULD AND AN ELECTRODE
 DURING MELTING OF THE ELECTRODE, SUCH RELATIVE MOTION BEING
 TRANSVERSE TO THE VERTICAL AXIS OF THE MOULD AND BEING EFFECTED
 WITHOUT ROTATION OF THE MOULD ABOUT ITS VERTICAL AXIS.

UK      01203532
   METALLGESELLSCHAFT A.G.
   669.181.25          669.181.422
      PROCESSES                 DIRECT REDUCTION    IRON OXIDE
      KILNS                     REDUCING AGENTS     ATMOSPHERES
      SPONG IRON                PRODUCTS            HEAT CONTENT
      EXHAUST GASES
          P01468  B18171           080268GE  30/01/69  26/08/70
 A PROCESS FOR DIRECTLY REDUCING MATERIALS CONTAINING IRON OXIDE
 IN A ROTARY KILN BY MEANS OF SOLID CARBONACEOUS REDUCING AGENTS,
 WHEREIN THE KILN CHARGE AND THE KILN ATMOSPHERE MOVE IN CO-CURRENT
 THROUGH THE KILN AND THE MATERIAL DISCHARGED FROM THE KILN IS
 COOLED AND SEPARATED TO PROVIDE A SPONGE IRON PRODUCT, AND WHEREIN
 THE HEAT CONTENT AT LEAST OF A PARTIAL STREAM OF THE EXHAUST GASES
 FROM THE KILN IS USED TO DRY AND/OR HEAT THE COOLED AND SEPARATED
 SPONGE IRON PRODUCT UNDER NON-OXIDIZING CONDITIONS BEFORE OR WHEN
 SAID PRODUCT IS SUBJECTED TO FURTHER PROCESSING.

UK      01203706
   METALLGESELLSCHAFT A.G.
   622.742             669.141.17            669.181.25
      PROCESSES                 MAGNETIZATION          DISCHARGE
      SOLIDS                    KILNS                  REDUCTION (GENERAL)
      IRON                      MATERIALS              SCREENING (SIZING)
      SPONGE IRON               CORE
          P01666  G181             230168GE  22/01/69  03/09/70
 A PROCESS FOR DRESSING THE SOLIDS MIXTURE DISCHRAGE FROM A ROTARY
 KILN USED FOR THE DIRECT REDUCTION OF OXIDIC, IRON-CONTAINING
 MATERIALS, WHEREIN THE SOLIDS MIXTURE WHICH IS DISCHARGED FOR THE
 ROTARY KILN IS COOLED AND SUBJECTED TO MULTI-STAGE SCREENING TO
 PROVIDE A PARTICLE SIZE FRACTION OF 1 TO 10 MILIMETRES, AND WHEREIN
 THIS FRACTION IS SUBJECTED TO A TWO STAGE MAGNETIC SEPARATION,
 WHICH IS OPERATED IN ONE STAGE AT A LOW FIELD STRENGTH BELOW 100
 OERSTEDS AND IN THE OTHER STAGE AT A HIGHER FIELD STRENGTH ABOVE
 1000 OERSTEDS IN ORDER TO SEPARATE THE MIXTURE INTO A SPONGE IRON
 FRACTION AND TWO COKE FRACTIONS ONE OF WHICH HAS A LOWER ASH
 CONTENT THAN THE OTHER, THE COKE FRACTION HAVING THE LOWER ASH
 SEPARATED OUT IN THE STAGE OPERATED AT THE LOW FIELD STRENGTH.

UK      0120[illegible]08
   GEBR. BOHLER AND CO. A.G.
   669.054.7
      IMPROVEMENT               ELECTROSLAG PROCESS COOLING
      MOULDS                    SOLIDIFICATION         LIQUID METALS
      INGOTS
          P01566  S1B181T1         280668AU  26/02/69  09/09/70
 AN ELECTRO-SLAG REFINING PROCESS, WHICH COMPRISES COOLING A MOULD
 HAVING AN OPEN LOWER END, SOLIDIFYING MOLTEN METAL IN THE MOULD
 TO FORM AN INGOT UNDER A POOL OF MOLTEN METAL, SUPPORTING THE
```

```
  INGOT AT ITS LOWER END ON A BOTTOM PLATE, AND INTERMITTENTLY
  LIFTING THE MOULD RELATIVE TO THE INGOT IN A PLURALITY OF STEPS,
  WHICH ARE CONTROLLED TO PREVENT A LIFTING OF THE LOWER END OF THE
  MOULD ABOVE THE LOWERMOST POINT OF THE POOL OF MOLTEN METAL.

UK      01206065
   ALLEGHENY LUDLUM STEEL CORP.
   669.054.7
        IMPROVEMENT          PROCESSES            INGOTS
        PRODUCTION           CONSUMABLE           ELECTRODES
        POWER                MOULDS               LIQUID METALS
        CATHODES             ATMOSPHERES          PRESSURE
        WELDING              ELECTRIC ARCS
            P01639  B1               160267US  14/02/68  23/09/70
  A PROCESS FOR PRODUCING A METAL INGOT BY THE CONSUMABLE ELECTRODE
  MELTING METHOD WHEREIN AN ELECTRODE OF THE METAL TO BE MELTED
  IS ASSEMBLED WITHIN A CONSUMABLE ELECTRODE FURNACE STRUCTURE WHICH
  IS CONNECTED TO A POWER SUPPLY AND WHICH INCLUDES A MOULD ADAPTED
  TO CONTAIN MOLTEN METAL, COMPRISING THE OPERATING OF SAID
  ELECTRODE AS THE ANODE AND THE MOULD AS THE CATHODE IN A
  NONREACTIVE GASEOUS ATMOSPHERE AT A SUBATMOSPHERIC PRESSURE OF
  FROM 30 MICRONS TO 2000 MICRONS, SAID PRESSURE SUBSTANTIALLY
  PRECLUDING MOLTEN METAL FROM WELDING TO THE SIDE OF SAID MOULD;
  ENERGIZING THE POWER SUPPLY TO MAINTAIN AN ELECTRICAL ARC TO
  MELT THE ELECTRODE; AND CONTINUING THE MELTING UNTIL THE ELECTRODE
  IS SUBSTANTIALLY CONSUMED.

UK      01207003
   DIRECTOR OF NATIONAL RESEARCH INSTITUTE OF METALS
   669.183.5            669.18-932
        APPARATUS            METHOD               REFINING
        LIQUID METALS        FURNACES             FLOW
        LANCES               OXYGEN               SLAGS
        BLOWING              EXHAUST GASES
            P01676  B1B1T1           140867JA  13/08/68  30/09/70
  APPARATUS FOR CONTINUOUS REFINING OF A MOLTEN METAL COMPRISING AT
  LEAST TWO TROUGH TYPE FURNACES ARRANGED IN SERIES SO THAT A MOLTEN
  METAL MAY BE CONDUCTED FROM ONE FURNACE TO THE NEXT FURNACE, EACH
  OF SAID FURNACES INCLUDING A FLOW INLET FOR A MOLTEN METAL AT ONE
  END, A FLOW OUTLET FOR AN OVERFLOW OF A REFINED MOLTEN METAL AT THE
  OPPOSITE END, A BLOWING ZONE PROVIDED WITH AT LEAST ONE LANCE
  CAPABLE OF FEEDING OXYGEN AND A SLAG-FORMING MATERIAL INTO SAID
  BLOWING ZONE, WHEREIN A MOLTEN METAL MAY BE VIGOROUSLY MIXED AND
  REACTED WITH SAID OXYGEN AND SLAG-FORMING MATERIAL AND A FLOW ZONE
  ADJACENT THERETO, WHEREIN THE MOLTEN METAL AND THE FORMED SLAG MAY
  BE REACTED WHILE FLOWING COUNTER-CURRENTLY AND THE SLAG MAY BE
  ZONES BEING SITUATED BETWEEN SAID INLET AND OUTLET.

UK      01209464
   FOSECO INTERNATIONAL LTD.
   669.054.7            621.791.793
        STARTING             WELDING              SLAGS
        ELECTROSLAG PROCESS  FIRING               SPRINGS
            P01818  B1               23/06/67            21/10/70
  A STARTER TABLET FOR USE IN THE ELECTROSLAG REMELTING OR WELDING
  PROCESS WHICH IS A BLOCK OF A COMPOSITION WHICH COMPRISES
  INGREDIENTS WHICH REACT TOGETHER EXOTHERMICALLY WHEN FIRED TO
  YIELD A MOLTEN SLAG WHEREIN THE BLOCK HAS WITHIN IT, AND PROJECTING
  FROM IT ON OPPOSITE FACES, METAL SPRINGS.
```

```
UK      01209197
   INSTITUT ELEKTROSVARKI IMENI E.O.
   669.054.7           669.054.82          621.3.035.2
   669.14-412
        PROCESSES           CASTING             MOULDS
        PRODUCTION          ELECTRODES          SLAGS
            P01824  T1B1G1S1      250267RU  26/02/68  21/10/70
  A PROCESS FOR THE PRODUCTION OF METAL CASTINGS UTILISING A MOULD
  WITH A STATIONARY CONSUMABLE ELECTRODE THEREIN INCLUDING THE STEPS
  OF INTRODUCING SLAG INTO THE MOULD IN SUCH AN AMOUNT TO ENSURE THE
  LOWER END OF THE STATIONARY CONSUMABLE ELECTRODE IS IMMERSED IN A
  SLAG POOL FORMED IN THE MOULD AND ADDING FURTHER SLAG INTO THE
  MOULD AS THE ELECTRODE MELTS TO ENSURE THE SAID LOWER END REMAINS
  IMMERSED IN THE SLAG.

UK      01209229
   INSTITUT ELEKTROVARKI IMENI E.O. PARONA
   669.054             661.938
        METHOD              IMPROVEMENT         ATMOSPHERES
        ELECTROSLAG PROCESS NITROGEN            GASES
        PRESSURE
            P01827  S2T1B1G1      11/06/68            21/10/70
  A METHOD OF ELECTROSLAG REMELTING OF METALS CARRIED OUT IN AN
  ATMOSPHERE OF NITROGEN OR A MIXTURE OF NITROGEN AND INERT GAS UNDER
  PRESSURE IN EXCESS OF ATMOSPHERIC PRESSURE.

UK      01210236
   BRITISH IRON AND STEEL RESEARCH ASSOCIATION
   669.054.7           661.939.3
        PROCESSES           CASTING             MOULDS
        REFINING            SLAGS               WATER
        COPPER              COOLING             REDUCTION
            P01878  B1G1T1S1      25/05/67  20/05/68  28/10/70
  IN A METHOD OF REFINING METALS A PLASMA TORCH IS DIRECTED ON TO THE
  SURFACE OF A SLAG WITHIN A WATER-COOLED COPPER MOULD TO HEAT THE
  SLAG. AT STARTUP, A STOOL SEALS THE LOWER MOULD OPENING AND MOLTEN
  SLAG IS POURED INTO THE MOULD. TO ASSIST THE FORMATION OF AN ARC
  A STARTER PEG IS PROVIDED ON THE BASE-PLATE OR AN ELECTRICALLY
  CONDUCTING MATERIAL IS ADDED TO THE SLAG. METAL TO BE REFINED IS
  INTRODUCED INTO THE SLAG AS BARS, RODS, SCRAP, SHAVINGS, WIRES OR
  POWER AND IS FUSED TO FORM AN INGOT THAT IS CONTINUOUSLY WITHDRAWN
  FROM THE MOULD. ALTERNATIVELY, MOLTEN METAL CAN BE POURED THROUGH
  THE SLAG OR POWDERED METAL CAN BE ENTRAINED IN THE PLASMA AND
  SPRAYED INTO THE SLAG. DIRECT REDUCTION OF A METAL COMPOUND TO A
  METAL MAY ALSO BE ACHIEVED USING A REDUCING SLAG (E.G. IRON
  BRIQUETTES CONTAINING 92% TOTAL IRON WITH 90% AS METALLIC IRON CAN
  BE REDUCED.

UK      01210567
   ROHLER AND CO. A.G. GBER.
   669.054.7           669.046.58
        SULPHUR             SLAGS               CASTING
        ELECTROSLAG PROCESS EXIDATION           ATMOSPHERES
            P01889  S1G1B1        28/03/67  28/03/68  28/10/70
  AN ELECTROSLAG REFINING PROCESS FOR STEEL INTENDED TO HAVE A
  HIGH SULPHUR CONTENT IS CHARACTERIZED IN THAT REMELTING IS
  IS EFFECTED WITH A SLAG WHICH HAS A SULPHUR CONTACT SUCH THAT
  THERE IS NO NET TRANSFER OF SULPHUR BETWEEN THE SLAG AND THE
  STEEL TO BE REMELTED AND OXIDATION OF SULPHUR IN THE SLAG
  IS PREVENTED BY PROVIDING A REDUCING ATMOSPHERE OVER THE
```

```
UK      01210802
    BRITISH IRON AND STEEL RESEARCH ASSOCIATION
    621.365.2             669.034.7              62-52
         ELECTRODES             INGOT MOULDS           MELTING
         ELECTROSLAG PROCESS INSULATION                ROTATION
         SENSORS
              P01903  S10101          07/02/67  05/02/68  02/11/70
  IN AN ELECTROSLAG REFINING PROCESS A PLURALITY OF CONSUMABLE
  ELECTRODES ARE MELTED INTO ONE OR MORE INGOT MOULDS AND A
  CHARACTERISTIC OF THE AMOUNT OR RATE OF MELTING IN RESPECT OF
  ONE OR MORE OF THE ELECTRODES IS DETECTED TO PRODUCE A CONTROL
  SIGNAL TO CONTROL TO CONTACT THE FEED OF ALL THE ELECTRODES
  IN COMMON AT THE SAME RATE OF FEED. THREE ELECTRODE CLAMPS
  ARE MOUNTED ON A COMMON VERTICALLY MOVABLE ELECTRODE HEAD
  BUT ARE INSULATED FROM IT BY INSULATORS. THREE CONSUMABLE
  ELECTRODES ARE HELD IN THE CLAMPS AND DEPEND INTO SEPARATE
  INGOT MOULDS THE ELECTRODES BEING MAINTAINED VERTICAL BY
  STEADY ROLLERS MOUNTED ON A BAR. EACH ELECTRODE IS CONNECTED
  TO ONE PHASE OF A THREE-PHASE SUPPLY THE RETURN CONDUCTOR
  OF WHICH IS CONNECTED TO THE BASES OF THE MOULDS. THE HEAD CAN
  BA RAISED OR LOWERED BY THE ROTATION OF A THREADED ROD.

UK      01209975
    ASSOCIATED LEAD MFRS. LTD.
    669.034.7
         ELECTROLYTES          CATHODS                ANODES
         LIQUID METALS         REFINING
              P01925  B1              10/04/68  02/07/69  28/10/70
  ELECTROLYTIC REFINING OF METALS USING A FUSED SALT ELECTROLYTE
  IS CARRIED OUT IN A CELL CONTAINING A LOWER MOLTEN METAL
  CATHODE AN UPPER MOLTEN METAL ANODE, A FLUX (ELECTROLYTE)
  IMPREGNATED POROUS DIAPHRAGM BENEATH AND IN CONTACT WITH THE
  ANODE, A LAYER OF MOLTEN FLUX (MOLTEN ELECTROLYTE) BETWEEN THE
  CATHODE AND THE UNDERSIDE OF THE DIAPHRAGM AND, AND ANOTHER
  LAYER OF MOLTEN FLUX ON TOP OF THE ANODE AND IN CONTACT WITH
  THE DIAPHRAGM. THE DIAPHRAGM MAY CONSIST OF A FLEXIBLE FELT OF
  CERAMIC FIBRES. THE CELL SHOWN COMPRISES A STEEL CASTING LINED
  WITH BLOCKS OF CERAMIC MATERIAL WHICH ARE SHAPED TO FORM A WELL
  IN WHICH IS CONTAINED THE CATHODE AND ELECTROLYTE LAYER. THE
  WELL BEING PROVIDED WITH TUBES WHICH PROVIDE A SYPHON
  ARRANGEMENT FOR THE WITHDRAWAL OF THE MOLTEN METAL DEPOSITED
  AT AND FORMING THE CATHODE.

UK      01211945
    VOLOKHONSKY L A     PELTS B S            ZABOLOTNY J V
    ARTAMONOV V I       IVANOV L M           ET AL
    621.365.2           669-982              669.18-932
         CRUCIBLES             MELTING                ELECTRODES
         VACUUM FURNACES       ELECTRIC HEATING       LIQUID METALS
         ELECTRON BEAMS
              P02018  B10101          22/02/68            11/11/70
  A VACUUM OR INERT GAS-FILLED FURNACE COMPRISES A TILTABLE
  METAL MELTING CRUCIBLE, HERMETICALLY CLOSED BY A COVER
  SUPPORTING A CONSUMABLE ELECTRODE OR OTHER ELECTRIC HEATING
  MEANS, AND ONE OR MORE POURING CHAMBERS HERMETICALLY CONNECTED
  TO THE MELTING CRUCIBLE. CHAMBERS ACCOMMODATE ROTARY TABLES
  WITH MOULDS INTO WHICH MOLTEN METAL IS POURED FROM THE
  POURING SPOUT OF CRUCIBLE VIA A POURING TROUGH, AROUND THE
```

FILMED FROM BEST AVAILABLE COPY

CRUCIBLE IS A WATER-COOLED JACKET AND A SOLENOID FOR STIRRING THE MOLTEN METAL. THE CRUCIBLE IS CONNECTED TO AN EXHAUSTING DEVICE OR TO A SOURCE OF PROTECTIVE GAS. THE CRUCIBLE IS MOUNTED ON TRUNNIONS FOR TURNING BY MEANS OF A DRIVE MECHANISM AND A GEAR TRAIN. INSTEAD OF A CONSUMABLE ELECTRODE, THE HEATING MAY BE BY AN ELECTRON-BEAM GUN OR A PLASMA TORCH AND A CHARGE FOR MELTING CAN BE LOADED INTO THE CRUCIBLE.

```
UK      01212400
   G28 BOHLER AND CO. A.G.
   621.744.3            669.056.7
       MOULDS               CASTING              APPARATUS
       SEALING              ELECTRODES           MOTORS
          P02062  B              06/11/67  06/11/68  18/11/70
```

A MOULD IS SUPPORTED BY A HOLDING DEVICE WHICH IS CONSTRUCTED AS A MOULD LIFTING CARRIAGE HAVING TWO ARMS IN FORK FORMATION THE ARMS HAVING CHANNELS THROUGH WHICH THE SUPPLY AND WITHDRAWAL OF A COOLING AGENT TO AND FROM THE MOULD TAKES PLACE AND THE MOULD HAVING SIDE EXTENSIONS WHICH REST ON THP ARMS. SEALING ELEMENTS SURROUND THE OPENINGS OF THE CHANNELS AND GUIDE PINS MAU BE PROVIDED ON THE ARMS TO HOLD THE MOULD IN THE CORRECT POSITION, THE MOULD AND HOLDING DEVICE MAY FORM PART OF AN ELECTROSLAG INSTALLATION WHICH ALSO INCLUDES AN ELECTRODE AND ELECTRODE HOLDER VERTICALLY MOVABLE BY MEANS OF A MOTUR IS VERTICALLY MOVABLE BY MOTOR AND CABLE.

Question 1 :

```
                                                              10/06.
                 NEW PROCESSES FOR PRODUCTION OF STEEL (KEYWORDS)   1
UK      01175982
   ASSOCIATED ELECTRICAL INDUSTRIES LTD.
   669.14-412            621.365.22            669.14-477
      INGOT CASTING         TUBES                 ELECTROSLAG PROCESS
      ELECTRODES
         P00001  8151          19/04/67               01/01/70
 A METHOD OF PRODUCING A TABULAR METAL INGOT COMPRISING THE STEPS OF
 SUSPENDING A PLURALITY OF ELECTRODES SO THAT THEY DEPEND INTO AN
 ANNULAR MOULD SPACE DEFINED BY INNER AND OUTER WALLS, AT SPACED
 LOCATIONS AROUND THE MOULD SPACE, INTRODUCING SLAG INTO THE MOULD
 SPACE TO A LEVEL SUCH AS TO IMMERSE THE LOWER ENDS OF THE
 ELECTRODES AND, WHILST THE WALLS ARE COOLED, UTILISING THE
 ELECTRODES FOR CARRYING OUT AN ELECTRO-SLAG MELTING PROCESS SO AS
 TO CAUSE THE ELECTRODES TO MELT AND FORM THE TUBULAR INGOT IN THE
 MOULD SPACE.

UK      01177387
   ASSOCIATED ELECTRICAL INDUSTRIES LTD.
   621.365.22            669.14-412
      IMPROVEMENT           INGOTS                PRODUCTION
      ELECTROSLAG PROCESS
         P00085  S15281        19/04/67               16/01/70
 A METHOD OF PRODUCING METAL INGOTS HAVING A RADIALLY VARYING
 COMPOSITIONAL ANALYSIS UTILISING AN ELECTROSLAG PROCESS WHEREIN A
 RELATIVELY LARGE CENTRAL ELECTRODE AND AT LEAST ONE SURROUNDING
 RING OF SMALLER ELECTRODES DEPEND INTO MOLTEN SLAG WITHIN A MOULD,
 THE RELATIVE COMPOSITIONS OF THE ELECTRODES BEING SUCH AS TO GIVE
 THE REQUIRED RADIALLY VARYING ANALYSIS IN THE FINAL INGOT, AND
 ALL OF THE ELECTRODES BEING PROGRESSIVELY FUSED OFF INTO A MOLTEN
 METAL POOL UNDER THE SLAG.

UK      01179958
   WELLMAN ENGINEERING CORP.
   669.18-932
      SPRAY REFINING        IRON                  FLOW MEASUREMENT
      LADLES                TUNDISH               CARBON
      SILICON               ANALYZING
         P00188  B1G1          20/06/67               06/02/70
 A METHOD OF MEASUREMENT OF IRON IN SPRAY STEELMAKING, COMPRISING
 MEASURING THE WEIGHT OF IRON IN AN IRON-RECEIVING LADLE, USING THE
 SAID WEIGHT MEASUREMENT IN RELATION TO A PREDETERMINED TIME
 INTERVAL TO PROVIDE AN INDICATION OF THE FLOW-RATE FROM THE SAID
 LADLE, MEASURING AND INDICATING THE HEIGHT OF METAL IN A TUNDISH
 OR OTHER RECEPTACLE INTERPORED BETWEEN THE IRON-RECEIVING LADLE
 AND THE SPRAY STEELMAKING UNIT, MEASURING THE TEMPERATURE OF THE
 METAL IN THE SAID TUNDISH OR OTHER RECEPTACLE, APPLYING THE SAID
 HEIGHT AND TEMPERATURE MEASUREMENTS TOGETHER WITH A SIGNAL BASED
 UPON THE KNOUN CARBON SILICON ANALYSIS OF THE IRON, TO AN
 INSTRUMENT WHICH PROVIDES AN OUTPUT SIGNAL PROPORTIONAL TO THE
 FLOW-RATE FROM THE SAID TUNDISH OR OTHER RECEPTACLE AND APPLYING
 SAID OUTPUT SIGNAL TO AN INDICATOR WHICH PROVIDES AN INDICATION OF
 THE SAID FLOW-RATE FROM THE TUNDISH OR OTHER RECEPTACLE IN FLOW
UNITS

UK      01181653
   FOSECO INTERNATIONAL LTD.
   621.365.22            669.046.5             669.14-413
      STEELS                ELECTRODES            MELTING
      INGOTS                ELECTROSLAG PROCESS EXOTHERMIC REACTION
```

```
     SLAGS                   SULPHUR                 REDUCTION(GENERAL)
          P00277  S181              05/05/66              18/02/70
  A METHOD FOR THE PRODUCTION OF INGOTS OF STEEL OR OTHER METAL BY
  THE ELECTROSLAG REMELTING PROCESS, IN WHICH AN ELECTRODE OF STEEL
  OR OTHER METAL IS EMPLOYED,WHEREIN PRIOR TO THE INITIATION OF
  REMELTING THERE IS LOCATED ADJACENT THE TIP OF THE ELECTRODE A
  COMPOSITION COMPRISING INGREDIENTS WHICH REACT EXTHERMICALLY WHEN
  FIRED AND WHICH BURN TO YIELD A SULPHUR-REDUCING SLAG,AND WHEREIN
  AFTER THE REMELTING HAS COMMENCED,ADDITIONAL SLAG FORMING
  INGREDIENTS ARE ADDED WHICH MELT TO MAINTAIN THE MOLTEN SLAG.

UK      01184278
   GEBR BOHLER AND CO. A.G.
   669,056,7
       INSTALLING            ELECTROSLAG PROCESS METALS
       CONSUMABLE            ELECTRODES          MOULDS
       LIQUID METALS         ELECTRIC CURRENT    FLOW
          P00404  S181              061167AU  06/11/68  11/03/70
  AN INSTALLATION FOR THE ELECTRO SLAG REMELTING OF METALS, HAVING
  A MAIN ELECTRODE CONSISTING OF CONSUMABLE METAL, A MOULD AND AN
  AUXILIARY ELECTRODE SERVING FOR MAINTAINING THE HEAT IN THE
  MOLTEN METAL FORMED, AND THROUGH WHICH CURRENT FLOWS DURING THE
  PROCESS OF THE EXCHANGING OF THE MAIN ELECTRODE, CHARACTERISED IN
  THAT THE AUXILIARY, AND/OR THE MOULD, AND THE MAIN ELECTRODE,
  ARE ARRANGED SO AS TO BE MOVABLE ON TRANSPORT DEVICES WHICH ARE
  SEPARATE FROM EACH OTHER.

UK      01175982
   ASSOCIATED ELECTRICAL INDUSTRIES LTD.
   669,14-462           669,14-412            669,056,7
       PRODUCTION            TUBULAR               METALS
       INGOTS                ELECTROSLAG PROCESS COOLING
       LIQUIDS               FLOW                  ELECTRODES
       MOULDS                ALTERNATING CURRENT
          P00416  T1                19/04/67  19/04/68  01/01/70
  TUBULAR METAL INGOTS ARE PRODUCED BY AN ELECTROSLAG PROCESS IN AN
  ANNULAR MOULD.  COOLING LIQUID FLOWS DOWN THE OUTER SURFACE OF THE
  OUTER WALL AND THE INNER SURFACE OF THE INNER WALL INTO AN ANNULAR
  TROUGH.  ELECTRODES AT SPACED LOCATIONS AROUND THE MOULD ARE
  SUPPORTED IN GROUPS OF THREE TO WHICH A THREE-PHASE A.C. CURRENT
  IS SUPPLIED .  A COOLING LIQUID DRAIN FROM THE CENTRAL REGION IS
  SURROUNDED BY A BARRIER TO FORM AN ANNULAR CHANNEL INTO WHICH
  COOLING LIQUID MAY BE INTRODUCED SEPARATELY.  A BARRIER WITHIN THE
  TROUGH FORMS AN ANNULAR CHANNEL WITH THE MOULD INTO WHICH THE
  COOLING LIQUID MAY BE INTRODUCED SEPARATELY.

UK      01187148
   ASSOCIATED ELECTRICAL INDUSTRIES LTD.
   669,046,587          621,746,534
       IMPROVEMENT           COOLING               MOULDS
       ELECTROSLAG PROCESS WATER                   FLOW
       WALLS                 LAYERS                OPENINGS
          P00636  S181              20/04/67              08/04/70
  A METHOD OF COOLING A MOULD USED IN AN ELECTROSLAG REFINING PROCESS,
  WITH THE AID OF AN UNCONFINED CURTAIN OF WATER WHICH IS CAUSED TO
  FLOW OVER THE OUTER SURFACE OF A PART OF THEMOULD, THE METHOD
  INCLUDING THE STEPS OF SURROUNDING THE MOULD, VERTICALLY ABOVE
  THE SAID SURFACE, WITH A FUNNELLIKE STRUCTURE HAVING AN ANNULAR
  INTERNAL SURFACE WHICH EXTENDS FROM AN UPPER-LOCATED PORTION OF
  THE STRUCTURE TO TERMINATE IN A LOWER-LOCATED LIP PORTION OF THE
  STRUCTURE, THE SAID LIP PORTION SO CO-OPERATING WITH THE WALL OF
  THE MOULD TO DEFINE THEREBETWEEN A UNIFORM OR APPROXIMATELY UNIFORM
  ANNULAR ORIFICE, AND SUPPLYING WATER TO SAID STRUCTURE AROUND AN
  UPPER PART OF SAID ANNULAR INTERNAL SURFACE TO CAUSE IT TO FLOW
  DOWNWARDS AS A LAYER OVER SAID SURFACE TO BE DISCHARGED FROM THE
  ORIFICE TOWARDS THE SURFACE OF THE MOULD, THE ORIFICE BEING OF
  SUFFICIENTLY LARGE ANNULAR WIDTH, AND THE SAID INTERNAL SURFACE
SO TAPERING IN DIAMETER TOWARDS THE SAID LIP POR

UK      01188028
   INSTITUT ELEKTROSVARKI IMENI E.O.
   669.046.58          621.791.042
       METHOD                 ELECTROSLAG PROCESS CONSUMABLE
       ELECTRODES             MOULDS                ELECTRIC CURRENT
       AMPLITUDES             METAL                 DROPS
           P00685  S181            28/06/67             15/06/70
  A METHOD OF ELECTROSLAG REMELTING OF CONSUMABLE ELECTRODES IN A
  COOLED MOULD, WHEREIN SHORT PULSES OF ELECTRIC CURRENT ARE
  SUPERIMPOSED ON THE WORKING CURRENT OF THE ELECTRODE, THE
  AMPLITUDE AND DURATION OF SAID PULSES BEING SUFFICIENT FOR
  EFFECTING POSITIVE BREAKING OFF OF THE ELECTRODE METAL DROPS
  DURING THE FORMATION OF THE DROP UP TO THE MOMENT OF ITS NATURAL
  BREAKING OFF.

UK      01190075
   ASSOCIATED ELECTRICAL INDUSTRIES LTD.
   669.054.7           669.14-412
       IMPROVEMENT            ELECTROSLAG PROCESS PLANTS
       INGOTS                 PRODUCTION            ELECTRODES
       MELTING                ELECTRIC CURRENT      VARIATIONS
           P00778  319181          07/06/67             29/04/70
  ELECTROSLAG PLANT FOR INGOT PRODUCTION FROM A PLURALITY OF PROCESS
  ELECTRODES CONNECTIBLE TO A COMMON ELECTRICAL SUPPLY FOR
  CONTEMPORANEOUS MELTING, THE PLANT BEING ARRANGED FOR CONTROL OF
  THE PROGRESSIVE FEEDING OF THE ELECTRODES DURING THE PROCESS BY
  A SINGLE CONTROLLER RESPONSIVE TO ELECTRODE CURRENT VARIATIONS.

UK      01194968
   GEBR. BOHLER AND CO. A.G.
   669.054.7           621.3.036.6
       APPARATUS              JOINING               LINES
       ELECTRICITY            ELECTRODES            PLATE
       INSTALLING             ELECTROSLAG PROCESS METALS
           P01076  B1              201167AU  20/11/68  17/06/70
  A DEVICE FOR CONNECTING ELECTRIC LINES TO ELECTRODES, BASE PLATES,
  OR SIMILAR CONDUCTING ELEMENTS, IN INSTALLATIONS FOR THE ELECTRO-
  SLAG REMELTING OF METALS, IN WHICH TWO PIVOT LEVERS ARE PROVIDED,
  MOUNTED OPPOSITE EACH OTHER AND ROTATABLE IN THE PLACE DEFINED BY
  THEM, CARRYING IN EACH CASE AT LEAST ONE SLAMP JAW, BETWEEN WHICH
  PIVOT LEVERS THERE IS CONNECTED A PISTON CO-OPERATING WITH A
  PNEUMATIC OR HYDRAULIC CYLINDER, BY MEANS OF WHICH THE CLAMP
  JAWS CAN BE PRESSED ON THE CONDUCTING ELEMENT WHICH IS TO BE
  ATTACHED AND CAN BE RELEASED FROM THE CONDUCTING ELEMENTS WHICH
  HAS BEEN ATTACHED, THE PISTON BEING CONNECTED TO ONE OF THE PIVOT
  LEVERS, AND THE CYLINDER BEING CONNECTED TO THE OTHER PIVOT LEVER.

UK      01199703
   METALLGESELLSCHAFT A.G.
   669.181.25          669.181.4
       PROCESSES              PRODUCTION            SPONGE IRON
```

```
                                                                     10/06
                    NEW PROCESSES FOR PRODUCTION OF STEEL (KEYWORDS)   6
      SULPHUR              IRON ORES                 DIRECT REDUCTION
      KILNS                SOLID STATE               CARBON ALLOYS
      REDUCING AGENTS
           P01279  81T161            090268GE  27/01/69  22/07/70
   A PROCESS FOR PRODUCING SPONGE IRON HAVING A LOW SULPHUR CONTENT,
   WHEREIN A MATERIAL CONTAINING IRON OXIDE IS DIRECTLY REDUCED IN
   A ROTARY KILN WITH THE AID OF A SOLID CARBONACEOUS REDUCING AGENT
   WHICH IS SUPPLIED INTO THE ROTARY KILN AT THE DISCHARGE END
   THEREOF, AND WITH THE AID OF A SULPHUR-COMBINING SUBSTANCE, AND
   WHEREIN AT LEAST PART OF THE SULPHUR-COMBINING SUBSTANCE WHICH IS
   REQUIRED IS CHARGED DIRECTLY INTO THE REDUCING ZONE OF THE
   ROTARY KILN AND THE TOTAL WEIGHT OF THE CHARGED SULPHUR-
   COMBINING SUBSTANCE IS AT MOST 4% OF THE WEIGHT OF THE CHARGED
   ORE.

UK       01200278
    BOHLER AND CO. A.G. GEB
    669.054.7           669.14-412
      IMPROVEMENT          INSTALLING                MANUFACTURING
      INGOTS               VARIATIONS                DIMENSIONS
      ELECTROSLAG PROCESS STEELS                     ELECTRODES
      HOLDERS              MOULDS                    CONSUMABLE
           P01298  S1B1              141267AU  13/12/68  29/07/70
   AN INSTALLATION FOR THE MANUFACTURE OF INGOTS OF DIFFERENT SIZES
   IN THE ELECTRO-SLAG REMELTING OF METALS, IN PARTICULAR OF STEELS,
   CHARACTERIZED IN THAT THERE ARE PROVIDED TWO ELECTRODE HOLDERS
   DISPLACEABLE OR PIVOTABLE IN THE HORIZONTAL DIRECTION AND TWO
   MOULDS, WHEREIN THE ELECTRODE HOLDERS IN SEQUENCE ALTERNATELY
   SUPPLY CONSUBABLE ELECTRODES TO ONE OF THE TWO MOULDS, AND ONE OF
   THESE ELECTRODE HOLDERS INSERTS COMSUMABLE ELECTRODES INTO THE
   OTHER OF THE TWO MOULDS.

UK       01202192
    BRITISH IRON AND STEEL RESEARCH ASSOCIATION
    669.054.7           621.3.035.2           621.744.3
      IMPROVEMENT          METHOD                    ELECTROSLAG PROCESS
      MOTION               MOULDS                    ELECTRODES
      MELTING              APPARATUS
           P01366  810181            13/09/66            12/08/70
   A METHOD OF ELECTRO-SLAG REFINING IN WHICH NON-RECIPROCATING
   RELATIVE MOTION IS EFFECTED BETWEEN A MOULD AND AN ELECTRODE
   DURING MELTING OF THE ELECTRODE, SUCH RELATIVE MOTION BEING
   TRANSVERSE TO THE VERTICAL AXIS OF THE MOULD AND BEING EFFECTED
   WITHOUT ROTATION OF THE MOULD ABOUT ITS VERTICAL AXIS.

UK       01203532
    METALLGESELLSCHAFT A.G.
    669.181.25          669.181.422
      PROCESSES            DIRECT REDUCTION    IRON OXIDE
      KILNS                REDUCING AGENTS     ATMOSPHERES
      SPONG IRON           PRODUCTS            HEAT CONTENT
      EXHAUST GASES
           P01448  810171            080268GE  30/01/69  26/08/70
   A PROCESS FOR DIRECTLY REDUCING MATERIALS CONTAINING IRON OXIDE
   IN A ROTARY KILN BY MEANS OF SOLID CARBONACEOUS REDUCING AGENTS,
   WHEREIN THE KILN CHARGE AND THE KILN ATMOSPHERE MOVE IN CO-CURRENT
   THROUGH THE KILN AND THE MATERIAL DISCHARGED FROM THE KILN IS
   COOLED AND SEPARATED TO PROVIDE A SPONGE IRON PRODUCT, AND WHEREIN
   THE HEAT CONTENT AT LEAST OF A PARTIAL STREAM OF THE EXHAUST GASES
   FROM THE KILN IS USED TO DRY AND/OR HEAT THE COOLED AND SEPARATED
```

```
  SPONGE IRON PRODUCT UNDER NON-OXIDIZING CONDITIONS BEFORE OR WHEN
  SAID PRODUCT IS SUBJECTED TO FURTHER PROCESSING.

UK      01205008
   GEBR. BOHLER AND CO. A.G.
   669.054.7
        IMPROVEMENT          ELECTROSLAG PROCESS COOLING
        MOULDS               SOLIDIFICATION      LIQUID METALS
        INGOTS
            P01546  S181G1T1       280668AU  26/02/69  09/09/70
  AN ELECTRO-SLAG REFINING PROCESS, WHICH COMPRISES COOLING A MOULD
  HAVING AN OPEN LOWER-END, SOLIDIFYING MOLTEN METAL IN THE MOULD
  TO FORM AN INGOT UNDER A POOL OF MOLTEN METAL, SUPPORTING THE
  INGOT AT ITS LOWER END ON A BOTTOM PLATE, AND INTERMITTENTLY
  LIFTING THE MOULD RELATIVE TO THE INGOT IN A PLURALITY OF STEPS,
  WHICH ARE CONTROLLED TO PREVENT A LIFTING OF THE LOWER END OF THE
  MOULD ABOVE THE LOWERMOST POINT OF THE POOL OF MOLTEN METAL.

UK      01209464
   FOSECO INTERNATIONAL LTD.
   669.054.7           621.791.793
        STARTING            WELDING                SLAGS
        ELECTROSLAG PROCESS FIRING                 SPRINGS
            P01818  B1              23/06/67             21/10/70
  A STARTER TABLET FOR USE IN THE ELECTROSLAG REMELTING OR WELDING
  PROCESS WHICH IS A BLOCK OF A COMPOSITION WHICH COMPRISES
  INGREDIENTS WHICH REACT TOGETHER EXOTHERMICALLY WHEN FIRED TO
  YIELD A MOLTEN SLAG WHEREIN THE BLOCK HAS WITHIN IT, AND PROJECTING
  FROM IT ON OPPOSITE FACES, METAL SPRINGS.

UK      01209229
   INSTITUT ELEKTROVARKI IMENI E.O. PATONA
   669.054.7           661.938
        METHOD              IMPROVEMENT            ATMOSPHERES
        ELECTROSLAG PROCESS NITROGEN               GASES
        PRESSURE
            P01827  S2T1B1G1       11/06/68              21/10/70
  A METHOD OF ELECTROSLAG REMELTING OF METALS CARRIED OUT IN AN
  ATMOSPHERE OF NITROGEN OR A MIXTURE OF NITROGEN AND INERT GAS UNDER
  PRESSURE IN EXCESS OF ATMOSPHERIC PRESSURE.

UK      01210567
   BOHLER AND CO. A.G. GEBR.
   669.054.7           669.046.58
        SULPHUR             SLAGS                  CASTING
        ELECTROSLAG PROCESS EXIDATION              ATMOSPHERES
            P01889  S1G1B1          28/03/67  28/03/68  28/10/70
  AN ELECTROSLAG REFINING PROCESS FOR STEEL INTENDED TO HAVE A
  HIGH SULPHUR CONTENT IS CHARACTERIZED IN THAT REMELTING IS
  IS EFFECTED WITH A SLAG WHICH HAS A SULPHUR CONTACT SUCH THAT
  THERE IS NO NET TRANSFER OF SULPHUR BETWEEN THE SLAG AND THE
  STEEL TO BE REMELTED AND OXIDATION OF SULPHUR IN THE SLAG
  IS PREVENTED BY PROVIDING A REDUCING ATMOSPHERE OVER THE
  SLAG.

UK      01210802
   BRITISH IRON AND STEEL RESEARCH ASSOCIATION
   621.365.2           669.054.7            62-52
        ELECTRODES          INGOT MOULDS           MELTING
        ELECTROSLAG PROCESS INSULATION             ROTATION
```

```
                                                                    10/06
                 NEW PROCESSES FOR PRODUCTION OF STEEL (KEYWORDS)   6
      SENSORS
          P01903  S16181          07/02/67  05/02/68  02/11/70
IN AN ELECTROSLAG REFINING PROCESS A PLURALITY OF CONSUMABLE
ELECTRODES ARE MELTED INTO ONE OR MORE INGOT MOULDS AND A
CHARACTERISTIC OF THE AMOUNT OR RATE OF MELTING IN RESPECT OF
ONE OR MORE OF THE ELECTRODES IS DETECTED TO PRODUCE A CONTROL
SIGNAL TO CONTROL TO CONTACT THE FEED OF ALL THE ELECTRODES
IN COMMON AT THE SAME RATE OF FEED. THREE ELECTRODE CLAMPS
ARE MOUNTED ON A COMMON VERTICALLY MOVABLE ELECTRODE HEAD
BUT ARE INSULATED FROM IT BY INSULATORS. THREE CONSUMABLE
ELECTRODES ARE HELD IN THE CLAMPS AND DEPEND INTO SEPARATE
INGOT MOULDS THE ELECTRODES BEING MAINTAINED VERTICAL BY
STEADY ROLLERS MOUNTED ON A BAR. EACH ELECTRODE IS CONNECTED
TO ONE PHASE OF A THREE-PHASE SUPPLY THE RETURN CONDUCTOR
OF WHICH IS CONNECTED TO THE BASES OF THE MOULDS. THE HEAD CAN
BA RAISED OR LOWERED BY THE ROTATION OF A THREADED ROD.
```

Question 2 :

```
                                                                          41
                                                                       10/06
                     OXYGEN DEGASSING OF MOLTEN STEEL IN LADLE (UDC)     1
UK       01212391
   OTANI K
   669.046.554
        CASTING              PROCESSES           CARBON
        LIQUID METALS        REDUCING ATMOSPHERE CARBON MONOXIDE
        HYDROGEN             MOULDS              LADLES.
           P02061  31818171      19/01/67  15/01/68  18/11/70
  IN A METAL CASTING PROCESS, A MIXTURE OF AMORPHOUS CARBON OR
  ORGANIC CARBON COMPOUND AND HYDROCARBON IS PLACED IN DIRECT SONTACT
  WITH THE MOLTEN METAL AT SOME STAGE DURING THE PROCESS TO PRODUCE
  A REDUCING ATMOSPHERE OF CARBON MONOXIDE AND HYDROGEN GASES. THE
  MIXTURE MAY CONTACT THE MOLTEN METAL IN A SHOOT LEADING FROM A
  FURNACE TO A LADLE IN THE LADLE ITSELF OR IN THE MOULD, BY PLACING
  THE MIXTURE IN THE SHOOT LADLE OR MOULD PRIOR TO POURING OR BY
  SPRINKLING THE MIXTURE ON THE SURFACE OF THE MOLTEN METAL. THE
  HYDROCARBON MATERIAL IS PREFERABLY A PETROLEUM MATERIAL SUCH AS
  KEROSENE AND CAN BE APPLIED IN A LIQUID, GREASE IN A LIQUID OR
  WAX FORM. THE CARBON OR CARBONACEOUS MATERIAL MAY BE BREEZE,
  GRAPHITE, CHARCOAL OR ORGANIC MATERIALS SUUCH AS STRAW, CHAFF,
  PASTEBOARD, DRY FIBER, SAWDUST, CARDBOARD MILLBOARD OR STRAW AND
  WOOD ASH.
```

Question 3 :

```
                                                                        10/06
                         BLAST FURNACE TUYERES DESIGN (UDC)                 1
JK      01178614
   KOLSCH-FOLZER-WERKE A.G.
   669,162,222           669.162.221.2
        BLAST FURNACES        HOT BLAST            PIPES
           P00136  S2813161      20/04/67                21/01/70
  A MULTI-PART BLAST PIPE FOR BLAST FURNACES COMPRISING FIRST AND
  SECOND COAXIAL PIPES, AN EXPANSION CASE AND A BLAST NOZZLE, THE
  FIRST PIPE BEING SECURED TO A HOT AIR CHAMBER AND PROJECTING
  FREELY WITH ALL ROUND CLEARANCE INTO ONE END OF THE SECOND PIPE
  THE OTHER END OF WHICH COMMUNICATES WITH THE BLAST NOZZLE, THE
  EXPANSION CASE CONNECTING THE OVERLAPPING ENDS OF THE FIRST AND
  SECOND PIPES WITH ONE END OF THE CASE BEING SECURED TO THE FIRST
  PIPE AND THE OTHER END OF THE CASE BEING SECURED TO THE SECOND
  PIPE.

UK      01199741
   YAWATA IRON AND STEEL CO. LTD.
   669,162,221
        APPARATUS               INSTALLING           REMOVAL
        TUYERES                 BLAST FURNACES       HOLDERS
        IMPACT
           P01280  B17181          060966JA  31/08/67  22/07/70

UK      01204871
   DNEPROPETROVSKY ORDENA LENINA METALLURGICHESKY ZAVOD IM. PET
   ROVSKOGO
   669,162,221         62-82
        APPARATUS               COOLING              COOLANTS
        TUYERES                 BLAST FURNACES       VESSELS
        FLOW
           P01543  B181            11/09/67              09/09/70
  A DEVICE FOR REDUCING LOCAL OVERHEATING IN LIQUID COOLED MEMBERS
  EXPOSED TO HEAT FOR EXAMPLE TUYERES OF BLAST FURNACES, COMPRISING
  A VESSEL FOR HOLTING COOLANT LIQUID HAVING FLOW CONNECTIONS TO THE
  LIQUID COOLED MEMBER AND HAVING INLET MEANS AND OUTLET MEANS IN THE
  VESSEL SITUATED ON AXES WHICH INTERSECT WITHIN THE VESSEL AND TWO
  ELECTRODES IN THE VESSEL SEPARATED BY A GAP WHICH CO-INCIDES WITH
  THE INTERSECTION OF THE AXES FOR CREATING SHOCK WAVES IN THE
  LIQUID THEREBY INCREASING THE RATE OF FLOW THROUGH THE COOLED
  MEMBER.

UK      01212514
   PEEL E
   669,162,221.2       669.3
        TUYERES                 COPPER               WATER
        VALVES
           P02065  S1055171        13/07/67  01/04/68  18/11/70
  A CAST COPPER TUYERE SUBDIVIDED BY AN ANNULAR PARTITION INTO A
  NOSE SECTION AND A REAR PORTION  IS PROVIDED WITH INDEPENDENT WATER
  CIRCULATION MEANS FOR THE TWO PORTIONS. THE REAR PORTION HAS A
  WATER INLET AND OUTLET WHILST THE NOSE SECTION HAS A WATER INLET
  PIPE AND A WATER OUTLET PIPE WHICH DISCHARGES INTO A REAR PORTION
  AND IS PROVIDED WITH A NON-RETURN VALVE. IN A MODIFICATION PIPE
  EXTENDS REARWARDLY FROM PORTION THROUGH A SEPARATE DISCHARGE OUTLET
  THE TUYERE MAY BE MADE FROM A SINGLE OR FROM TWO INTERFITTING
  CASTINGS CONNECTED BY BOLTS.
```

Question 4 :

```
                                                                  10/06
                UTILIZATION OF BLAST FURNACE SLAG (UDC)              1
UK      01201305
   NATIONAL RESEARCH DEVELOPMENT CORP.
   669.162.275.2      669.054.82         666.1
        IMPROVEMENT        PROCESSES            PREPARATION
        SHAPE              OILS                 SOAPS
        CARBON COMPOUNDS   FLOUICLES            GRANULAR MATERIALS
        OXIDES             MIXTURES             NUCLEATION
        CRYSTALLIZATION    HEATING              GLASS
            P01334  B1G1         18/08/68             05/08/70
 A PROCESS FOR PREPARING A SHAPED GLASS CERAMIC ARTICLE WHICH
 COMPRISES HOT SHAPING IN A SHAPING MEANS A PARTICULATE GLASS
 BATCH FORMED FROM AN OXIDE MIX WITH A ZACHARIASEN N VALUE AS HEREIN
 DEFINED IN THE RANGE 0,28 - 0.4, IN WHICH NUCLEATION AND
 CRYSTALLISATION ARE EFFECTED DURING HEATING OF THE GLASS UP TO
 SHAPING TEMPERATURE IN THE SHAPING MEANS WITHOUT INTERMEDIATE
 MATURATION.
```

Question 5 :

```
                                                                10/06
                     IMPROVED RAILS (UDC)                           1
UK      01209151
   CHEMETRON CORP.
   621,923               621,791,053           669,14-426
        METHOD                GRINDING              RAILS
        WELDED JOINTS         LONGITUDINAL          PRESSURE
        ABRASIVES             APPARATUS
            P01844  G1                  21/11/66  25/10/67  21/10/70
  A METHOD OF GRINDING WELDS PRODUCED BY WELDING RAIL SECTIONS
  TOGETHER END-TO-END COMPRISES MOVING A TRAVELLING ABRASIVE SURFACE
  LONGITUDINALLY OF THE RAIL ACROSS THE WELD SUCCESSIVELY FROM
  OPPOSITE SIDES THEREOF AND FORCING THE ABRASIVE SURFACE INTO
  CONTACT WITH THE WELD AREA DURING THE LONGITUDINAL MOVEMENT IN
  SUCH A MANNER AS TO INCREASE THE GRINDING PRESSURE AGAINST THE RAIL
  IN THE AREA OF THE WELD AND TO REDUCE THE PRESSURE AS THE ABRASIVE
  SURFACE MOVES AWAY FROM THE WELD AREA. THE APPARATUS FOR CARRYING
  OUT THE METHOD MAY BE MOUNTED ON THE FLOOR OF A RAILWAY WAGGON
  AND CONSISTS OF A FRAME PROVIDED WITH A BASE HAVING WHEELS ENGAGING
  TRACKS IN THE FLOOR, ONE OF THE WHEELS BEING DRIVEN BY A MOTOR.
```

Question 5 :

```
                IMPROVED RAILS (KEYWORDS)
UK      01163794
   KLOCKNER-WERKE A.G.
   621.783.223.2        621.785.3
       IMPROVEMENT          ANNEALING FURNACES  CONTACTING
       PLATE                COOLANTS            RAILS
          P00387  T10181         2811676E  04/09/68  11/03/70
  AN ANNEALING FURNACE FOR THE SELECTIVE ANNEALING OF GROUND CONTACT
  PLATES HAVING ONE OR MORE RIDGES, WHICH INCLUDES COOLABLE RAILS
  EXTENDING ALONG THE DIRECTION IN WHICH THE GROUND CONTACT PLATES
  ARE ADAPTED TO TRAVEL THROUGH THE FURNACE AND INTO WHICH RAILS
  PASSING GROUND CONTACT PLATE RIDGES ARE ABLE TO PROJECT, THE RAILS
  BEING HOLLOW AND ADAPTED TO BE COOLED BY PASSING A COOLANT
  THERETHROUGH.
```

Question 6 :

```
                                                                      10/06.
                    EXTERNAL DESULPHURIZATION OF IRON (KEYWORDS)         1
UK      01205166
    NIPPON OIL CO. LTD.
    662.753.32              66.046.546.2
        PROCESSES               CRUDE OIL               OILS
        DESULPHURIZING          GASES                   HYDROGEN
        CATALYSIS               FLUIDIZING
            P01602  61                  12/06/67  12/06/68  16/09/70
  A PROCESS OF HYDRODESULPHURIZING HEAVY PETROLEUM HYDROCARBON OIL
  COMPRISES CONTACTING SAID OIL AND A HYDROGEN-CONTAINING GAS WITH
  A HYDRODESULPHURIZATION CATALYST IN THE FLUIDIZED STATE, IN WHICH
  PROCESS (A) SAID HEAVY OIL AND HYDROGEN-CONTAINING GAS ARE
  INTRODUCED INTO A REACTOR CONTAINING SAID CATALYST; (B) THE REACTION
  SYSTEM CONSISTING OF SAID HEAVY OIL,, HYDROGEN-CONTAINING GAS AND
  SAID CATALYST IS INTRODUCED TO A FLUIDIZATION INITIATING ZONE
  LOCATED IN THE REACTOR AT THE BOTTOM THEREOF TO INITIATE THE
  FLUIDIZATION OF THE SYSTEM; (C) THE RESULTING FLUIDIZED SYSTEM IS
  CAUSED TO MOVE FROM SAID FLUIDIZATION INITIATING ZONE INTO AN
  ADJOINING FLUIDIZED REACTION ZONE DISPOSED IN SAID REACTOR ABOVE THE
  FORMER ZONE, AND WHILE IT IS MOVED UPWARDLY THROUGH THE FLUIDIZED
  REACTION ZONE SUBSTANTIALLY ALL THE HYDRODESULPHURIZATION OF THE
  HEAVY OIL IS CARRIED OUT.

UK      01205481
    CHEVRON RUSEACH CO.
    662.753.32              66.046.546.2
        DESULPHURIZING          PROCESSES               PRODUCTION
        CRUDE OIL               SULPHUR                 CATALYSIS
        TEMPERATURE             BOILING
            P01604  61                  10/03/67  06/03/68  16/09/70
  A PROCESS FOR THE PRODUCTION OF LOW SULPHUR CONTENT HEAVY FUEL
  OIL CONTAINING NOT MORE THAN 2 WEIGHT PER CENT SULPHUR FROM HIGH
  SULPHUR CONTENT PETROLEUM RESIDUA CONTAINING MORE THAN 2 WEIGHT
  PER CENT OF SULPHUR COMPRISES SEPARATING HIGH SULPHUR PETROLEUM
  RESIDUE BOILING AT 650 F, OR ABOVE INTO AN ASPHALTIC FRACTION AND
  A NON-ASPHALTIC FRACTION, SUBJECTING THE NON-ASPHA[illegible] FRACTION TO
  A CATALYTIC HYDRODESULPHURIZATION TREATMENT, WHER[illegible]HE
  DESULPHURIZATION OF THE NON-ASPHALTIC FRACTION IS [illegible]RRIED OUT IN
  ADMIXTURE WITH LIGHT DISTILLATE BOILING AT A TEMPERATURE OF FROM
  300 F TO 700 F AND AT A TEMPERATURE OF FROM 300 F TO 700 F, AND AT
  LEAST PARTLY IN THE KEROSENE BOILING RANGE SAID DISTILLATE BEING
  PRESENT IN AMOUNT EFFECTIVE SUBSTANTIALLY TO INHIBIT HYDROCRACKING
  OF SAID NON-ASPHALTIC FRACTION.

UK      01210863
    TOPSOE H.F.A.
    66.074                  628.52
        GASES                   HYDROCARBONS            COMBUSTION
        DESULPHURIZING
            P01916  61                  10/02/67  09/02/68
  INDUSTRIAL GASES INCLUDING GASEOUS AND EVAPORATED HYDROCARBONS
  E.G. TOWN GAS, NATURAL GAS, BOTTLED GAS OR VAPOURIZED NAPHTHA
  ARE DESULPHURIZED BY PASSING THE GAS OR GAS MIXTURE THROUGH
  A BED OF ACTIVATED ZINC OXIDE.

UK      01212325
    DEUTSCHE EDELSTAHLWERKE A.G.
    669.187.25              669.046.546.2           669.046.517
    669-982
        ALLOY STEELS            REFINING                ELECTRIC FURNACES
```

10/06/
2

EXTERNAL DESULPHURIZATION OF IRON (KEYWORDS)

DESULPHURIZING   SLAGGING   SLAG
DEGASSING   OXYGEN   VACUUM
LADLES

P02048   S1G1T181   26/09/68   14/05/69   11/11/70

ALLOY STEELS ARE REFINED BY; A) PRODUCING A MELT OF ABOUT THE CORRECT COMPOSITION IN AN ELECTRIC ARC FURNACE; B) PARTIALLY DESULPHURIZING THE MELT IN A LADLE USING A SLAGGING AGENT (E.G. LIME-FLUORSPAR ASSOCIATED WITH ONE OR MORE OF A1, MG AND SI) C) REMOVING THE SLAG; D) CONTINUOUSLY SUBJECTING THE MELT TO VACUUM CIRCULATORY-DEGASSIFICATION WHILE INTRODUCING OXYGEN INTO THE LADLE AND A PULVERIZED SLAGGING AND DESULPHURIZING AGENT INTO THE LADLE; AND E) DESLAGGING THE MELT AND POURING OFF REFINED ALOOY STEEL. ALLOVING OR CONVENTIONAL DEOXIDIZING AGENTS ADDITIONS CAN BE MADE DURING THE VACUUM DEGASSIFICATION STEP.

Question 7 :

```
                                                                       10/06
                   IMPROVEMENTS IN TRANSPORTATION OF HOT METAL (K)       1
UK      01179976
   DEMAG A.G.
   621.746.222
        LIQUID METALS        TRANSPORTATION       CONTAINERS
        LININGS              CHARGING             OPENINGS
             P00198  81G151          041267GE  28/08/68  04/02/70
 A LINED, ELONGATED TRANSPORTATION CONTAINER OR VESSEL FOR LIQUID
 METALS, WHICH IS PROVIDED BETWEEN ITS ENDS WITH A CHARGING OPENING
 AND WHICH IS CARRIED ON BOGIES THROUGH THE INTERMEDIARY OF
 SUPPORTING TRUNNIONS WHICH ARE SECURED TO ITS ENDS, INCLUDING
 DETACHABLE CLOSURE PLATES WHICH ARE PROVIDED AT THE ENDS OF THE
 TRANSPORTATION VESSEL OR CONTAINER.

UK      01191222
   CANADIAN PATENTS AND DEVELOPMENT LTD.
   543.27              669.14-154           669.785
        APPARATUS            DETERMINATION        LIQUID METALS
        GASES                MEASUREMENT          INSTRUMENTATION
        THERMAL RESISTANCE   TUBES                ELECTROLYTES
        IONS                 TRANSPORTATION       ADHESIVES
             900855  71B161          150967CA  15/08/67  13/05/70
 A TIP FOR A PROBE FOR THE DIRECT DETERMINATION OF A GAS IN A MOLTEN
 METAL, THE PROBE TIP COMPRISING;
 (A) A HEAT-RESISTANT, ELECTRICALLY NON-CONDUCTIVE TUBE;
 (B) A SOLID MASS OF A MIXTURE OF A SELF-SINTERABLE SOLID ELECTROLYTE
 WHICH RETAINS ITS IONIC TRANSPORT PROPERTIES BUT WHICH DOES NOT
 EXHIBIT SUBSTANTIAL ELECTRONIC TRANSPORT PROPERTIES AT THE TEMP. OF
 THE  MOLTEN METAL AND A SUITABLE FUGITIVE BINDER DISPOSED AND
 SECURED AT ONE OPEN END OF THE HEAT-RESISTANT TUBE ; AND
 (C) AT LEAST ONE ELECTRICALLY CONDUCTIVE ELEMENT IN ELECTRICAL
 CONTACT WITH THE ELECTROLYTE.
```

Question 9 :

```
                                                                  10/06
                 IMPROVED BLAST FURNACE STOVE DESIGN (UDC)            1
UK      01176526
   NIPPON KOKAN KABUSHIKI KAISHA
   669.162.23
       STOVES                 HOT BLAST
            P00051  S1B161           60466JA  14/04/67  07/01/70
 A HOT BLAST STOVE, COMPRISING A COMBUSTION CHAMBER, A HEAT
 ACCUMULATING CHAMBER, A PARTITION WALL BETWEEN SAID CHAMBERS, A
 DOME AT THE TOP OF AND INTER-CONNECTING SAID COMBUSTION CHAMBER
 AND SAID HEAT ACCUMULATING CHAMBER, AND A PILE OF MEMBERS OF
 A REFRACTORY MATERIAL HAVING HAPHAZARD PASSAGES BETWEEN SAID
 MEMBERS, SAID PILE BEING POSITIONED AT THE TOP OF SAID HEAT
 ACCUMULATING CHAMBER AND WITHIN SAID DOME, SAID DOME HAVING A
 DOWNWARDLY CONCAVE INNER SURFACE AND SAID PILE GRADUALLY
 INCREASING IN DEPTH IN A DIRECTION FROM SAID PARTITION WALL AWAY
 FROM SAID COMBUSTION CHAMBER.

UK      01183154
   ZIMMERMANN AND JANSEN G.M.B.H.
   669.162.23
       HOT BLAST             STOVES               FUELS
       GASES                 BURNERS              AIR
       COMBUSTION            DUCTS                COOLANTS
       CIRCULATION           HOUSINGS
            P00355  S1B1            03/10/67  03/10/68  04/03/70
 IN A HOT BLAST STOVE IN WHICH A BURNER SUPPLIES SEPARATE STREAMS
 OF AIR AND GAS TO THE STOVE COMBUSTION CHAMBER MEANS FOR CLOSING
 OFF THE BURNER FROM THE COMBUSTION SPACE WHILST THE STOVE IS ON
 BLAST COMPRISES AN AXIALLY DISPLACEABLE CONE.   THE WALL OF THE
 CONE MAY HAVE DUCTS FORMED THEREIN FOR A CIRCULATING COOLANT.
 AS SHOWN THE BURNER COMPRISES A TUBULAR HOUSING SUBDIVIDED BY A
 FIXED TUBULAR PARTITION INTO CHAMBERS FOR AIR AND FUEL GAS
 RESPECTIVELY.   OVER THE OUTSIDE OF THE PARTITION SLIDES A SLEEVE
 CARRIED BY THE CONE AND ON THE INSIDE OF THE PARTITION MOVES A
 SLIDE MOVABLE INDEPENDENTLY OF THE CONE. BY MEANS SUCH AS A
 CROSSHEAD.   IN THE BURNING POSITION INDICATED AIR FLOWS FROM AN
 INLET AND CHAMBER TO A MIXING CHAMBER AND FUEL GAS FLOWS FROM AN
 INLET TO THE MIXING CHAMBER VIA THE CHAMBER.   REGULATION OF THE
 BURNER IN DEPENDENCE OF THE QUALITY OF THE FUEL GAS IS EFFECTED
BY MOVING THE SLIDE AND CONE.

UK      01203928
   KONINKLIJKE NEDERLANDSCHE HOOGOVENS EN STAALFABRIEKEN N.V.
   669.162.23
       METHOD                    GAS HEATING          STOVES
       AIR                       BLAST FURNACES       TEMPERATURE
            P01477  G1B1T1            070668NL  15/05/67  03/09/70
 A METHOD OF GAS HEATING UP A HOT BLAST STOVE FOR REGENERATIVELY
 HEATING THE AIR TO BE FED TO A BLAST FURNACE, CHARACTERISED IN
 THAT FLAME TEMPERATURE IN THE HOT BLAST STOVE SUBSTANTIALLY UPON THE
 BEGINNING OF HEATING UP IS CHOSEN TO BE SO MUCH HIGHER THAN THE AIR
 TEMPERATURE WHEN LEAVING THE BLAST STOVE, OR THE DOME TEMPERATURE
 IN THE BLAST STOVE AT THE END OF THE PRECEDING AIR BLAST PERIOD,
 AS TO CORRESPOND TO A VALUE 10% TO 40% OF THE TEMPERATURE
 DIFFERENCE BETWEEN THE MAXIMUM ADMISSIBLE FLAME TEMPERATURE AND
 AND SAID AIR OR DOME TEMPERATURE AT THE END OF THE PRECEDING BLAST
 PERIOD, AND IS INCREASED, GRADUALLY OR IN STEPS, UNTIL THE
 MAXIMUM ADMISSIBLE FLAME TEMPERATURE IS REACHED.
```

Question 9 :

```
                                                              10/06/
                  IMPROVED BLAST FURNACE STOVE DESIGN (KEYWORDS)      1
.JK     01203928
   KONINKLIJKE NEDERLANDSCHE HOOGOVENS EN STAALFABRIEKEN N.V.
   669.162.23
       METHOD                    GAS HEATING          STOVES
       AIR                       BLAST FURNACES       TEMPERATURE
            P01477  618171           070668NL  19/05/67  03/09/70
  A METHOD OF GAS HEATING UP A HOT BLAST STOVE FOR REGENERATIVELY
  HEATING THE AIR TO BE FED TO A BLAST FURNACE, CHARACTERISED IN
  THAT FLAME TEMPERATURE IN THE HOT BLAST STOVE SUBSTANTIALLY UPON THE
  BEGINNING OF HEATING UP IS CHOSEN TO BE SO MUCH HIGHER THAN THE AIR
  TEMPERATURE WHEN LEAVING THE BLAST STOVE, OR THE DOME TEMPERATURE
  IN THE BLAST STOVE AT THE END OF THE PRECEDING AIR BLAST PERIOD,
  AS TO CORRESPOND TO A VALUE 10% TO 40% OF THE TEMPERATURE
  DIFFERENCE BETWEEN THE MAXIMUM ADMISSIBLE FLAME TEMPERATURE AND
  AND SAID AIR OR DOME TEMPERATURE AT THE END OF THE PRECEDING BLAST
  PERIOD, AND IS INCREASED, GRADUALLY OR IN STEPS, UNTIL THE
  MAXIMUM ADMISSIBLE FLAME TEMPERATURE IS REACHED.
```

Question 10 :

```
                                                                        51
                                                                     10/06
                 REFRACTORY COATING OF OXYGEN LANCES (MDC)              1
UK      01193769
   DRESSER INDUSTRIES INC.
   669.046.564.5       666.76              669.183.212.6
       LANCES               COATINGS             REFRACTORIES
       STEELMAKING          ELONGATION           TUBES
       LAYERS               AGGREGATES           METAL FOILS
       MELTING POINTS       ALUMINIUM
            P01005  01710161     001167US  08/11/68  03/06/70
 A REFRACTORY-COATED LANCE SUITABLE FOR USE IN STEELMAKING, WHICH
 COMPRISES AN ELONGATED METAL TUBE HAVING ON AT LEAST A MAJOR PART
 OF ITS OUTER CYLINDRICAL SURFACE A PLURALITY OF LAYERS OF
 MOISTURE-FREE CHEMICALLY BONDED REFRACTORY AGGREGATE THE
 REFRACTORY LAYERS BEING SEPARATED BY METAL FOIL AND THE OUTER
 REFRACTORY LAYER BEING COVERED BY METAL FOIL, ALL THE FOIL
 HAVING A MELTING POINT EQUAL TO OR GREATER THAN THAT OF
 ALUMINIUM .
```

```
                REFRACTORY COATING OF OXYGEN LANCES (KEYWORDS)        10/06
                                                                          1
UK     01193769
   DRESSER INDUSTRIES INC.
   669,046,564.5      666.76                669,183,212.6
       LANCES              COATINGS              REFRACTORIES
       STEELMAKING         ELONGATION            TUBES
       LAYERS              AGGREGATES            METAL FOILS
       MELTING POINTS      ALUMINIUM
          P01005  S1T18161      081167US  08/11/68  03/06/70
 A REFRACTORY-COATED LANCE SUITABLE FOR USE IN STEELMAKING, WHICH
 COMPRISES AN ELONGATED METAL TUBE HAVING ON AT LEAST A MAJOR PART
 OF ITS OUTER CYLINDRICAL SURFACE A PLURALITY OF LAYERS OF
 MOISTURE-FREE CHEMICALLY BONDED REFRACTORY AGGREGATE THE
 REFRACTORY LAYERS BEING SEPARATED BY METAL FOIL AND THE OUTER
 REFRACTORY LAYER BEING COVERED BY METAL FOIL, ALL THE FOIL
 HAVING A MELTING POINT EQUAL TO OR GREATER THAN THAT OF
 ALUMINIUM.

UK     01195936
   FOSECO INTERNATIONAL LTD.
   669,046,564.5      666,764.6             666,763.46
       OXYGEN LANCES       PROTECTIVE COATINGS REFRACTORIES
       SLAGS               SOLID STATE           METALS
       STRIPS              FIBRES                CHROMITE REFRACTORS
       GRAPHITE            ALUMINIU,M            SILICA
       WATER
          P01124  S1T16181      11/06/68  03/06/69  26/06/70
 A LANCE HAS A PROTECTIVE LAYER OF COMPRESSIBLE REFRACTORY MATERIAL
 APPLIED TO THE END WHICH IS IMMERSED IN MOLTEN METAL, SO THAT
 SOLIDIFIED METAL OR SLAG MAY READILY BE REMOVED WITHOUT DAMAGING
 THE LANCE, SEMI-CYLINDRICAL SLEEVES OR STRIPS OF FIBRE-CONTAINING
 BONDED REFRACTORY ARE SECURED TO THE LANCE HEAD BY CLIPS, OR ARE
 STUCK TO THE LANCE HEAD. THE REFRACTORY LAYER AND ANY EXPOSED
 PARTS OF THE LANCE ARE COATED WITH A REFRACTORY NON-WET COATING
 SUCH AS A MIXTURE OF CHROMITE POWDER, GRAPHITE POWDER AND
 BENTONITE IN AN AQUEOUS SOLUTION OF ALUMINIUM ORTHO-PHOSPHATE, OR
 CHROMATE FLOUR WITH A SILICA SOL AND POLYSACCHARIDE GUM IN WATER.

UK     01212371
   MESSER GRIESHEIM G.M.B.H.
   621.74.047.002.5   669,046,517
       METHOD              DEGASSING             PURIFICATION
       CONTINUOUS CARTING  GASES                 LIQUID METALS
       JETS                MOULDS                LANCES
       REFRACTORIES        NOZZLES
          P02050                12/11/66  10/11/47  18/11/70
```

THE RELATION
OF THE UDC TO THE EURATOM THESAURUS FROM THE POINT OF VIEW OF A SMALL INFORMATION SERVICE

Aviva KARA/MAROSI/
Israel Atomic Commission,Soreq Nuclear Research,Library and
Technical Centre,Yavne,Israel

The Big Centres and their descriptor level.

Each Big Information Centre operates through upper, lower and side organizations. Its guiding principles and working methods are geared to the *general needs* of the whole information seeking community of its thematic field, thus it *is rather bibliographically oriented*.

Regarding subject organization, its basic approach is directed towards the achievement of *high recall* ratios through *high outputs*. Thus the *descriptor level* of the Big Centres' Thesaurus is rather *high*. - (The subject management of several large organizations which are based on the KWIC/KWOC indexing technique drops out of the frame of this consideration.) - The keywords and their specificity level, their connections are worked out on the basis of quantitative considerations, drawn directly from practice. The *structure of its vocabulary*, in contrast to the classical hierarchic systems is more up to date and *elastic*.

The need of the Small Information Service for lower level, hierarchical fixed dictionary

The purpose of a Small Service is to provide information for a small or medium sized community of users. The central task is to satisfy the *specific requirements* of its individual users.

The Small Service in part provides the same service as the Big Centre, but the two are not identical. The feature on which the subject organization of the Small Service is built is, that the *specificity of descriptors are already taken into account during the input, for the retrieval should not operate on high recall and low precision*.

The working methods of these two kinds of institutions are therefore reversed. To obtain a higher precision the Big Centre has to tighten the subject-area horizontally during the search process, while the Small Service has to narrow the generic level of the subject field in the input phase. It is advisable to work with a larger vocabulary stock at a lower generic level.

It means that the requirements of the Small Service cannot be met satisfactorily by the Big Centre's Thesauri only. It must have a well-elaborated hierarchic chain too, for the stabilization of the separate terms. On the other hand these stabilized terms must have a structure into which they can fit to form more complex units on a higher level than the original constituent parts.

These general reflections arised during the study-period of an experimental project at the Documentation Service of the Israel Atomic Commission at Nahal Soreq. It was aimed to prepare an alphabetical subject index for the book-section (20,000 units, classified according to the UDC) with the purpose that this index should serve as a basis for the comprehensive subject organization of the entire collection later on, (10,000 volumes back periodicals, - some 800 subscriptions annually, - 130,000 reports). For this material the subject access was via several outside sources.

By closer examination it became clear, that the construction of a self-conceived independent local subject index - (on the needed lower generic level) - in any Small Service is an almost impossible task. Neither would the simpler approach: the compilation of an index based on the higher levelled descriptos taken without change from the Big Centre's stock solve the problem.

These circumstances led to the conclusion, that by a much more radical approach, namely by constructing an index-language brought to a common denominator, containing a structurally unified, but multi-level vocabulary, which would embrace both the large Big Centre's and the narrower Small Service's terms could offer a chance for the solution.

Such a project would require international co-operation. Then, if the mechanized system of the Big Centre included the stock of lower level dictionary, and the information compiled - (including bibliographical and subject-descriptive data for each document) - could be centrally disseminated; the preparation of any local subject index could be reduced to reproducing and filing operations.

The problematics of the UDC/Euratom connection.

The aim of this discussion is to show, how a correlative coordinate index based on the vocabulary of the EURATOM Thesaurus supplemented with the UDC can perform the above mentioned task.

1/UDC

The UDC owing to its hierarchic structure offers an immediate sequence and at the same time a separate posting place for every member on every generic level of a given category of concept.

f.e.:	Animals:	Mammals:	Primates:	Apes:
	59	599	599.8	599.81

These features make it suitable for the planned stabilization of the term-level.

But not without difficulties; one of them is to be found in the ever growing correlative elements in the UDC, by which the leading concept of a given hierarchic-chain-unit is connected at the same generic level with other concept/s.

In the UDC language these correlative elements permit a main schedule term to be used as special auxiliary in the same group too, leading to considerable fragmentation.

for example:

Scattering, diffusion, diffraction of Elementary Particles: 539.12.172

but:

a/ Electron diffraction: $539.12\boxed{4}.172$

Neutron -"- $539.12\boxed{5.5}.172$

or, if the emphasis on the process itself:

b/ Scattering of electrons: 539.12.172 : $539.12\boxed{4}$!

Diffraction of neutrons: 539.12.172 : $539.12\boxed{5.5}$!

and another form - in different context:

c/ Neutron scattering (as Nuclear reactions)539.171.4

This classificational scattering is understandable. It is an inevitable effect of our scientific and technical progress on the nearly century-old system. Each newly emerging discipline obviously requires unified framing of its constituent parts; nevertheless because of the complexity of these new disciplines, the same elements can form distinct units by themselves too.

for example:

Health physics: 613.6 : 621.039

but some intrinsic elements of this entity are widely dispersed:

Radiation effects on human physiology 612.014.481

Radiation effects on particular parts of the body: 539.12.04 : 612.1...

Radiation injuries: 616-001.28

The hierarchic structure of the UDC does not allow the construction a well built cohesive unit for such interdependent components (i.e. correlative units), thus the dispersion cannot be avoided.

2/EURATOM Thesaurus

The strength of the Thesauri against the orthodox systems are due to their elaboration of the correlative-units.

In view of the active part played by the EURATOM in fostering international co-operation in the field of nuclear information, its Thesaurus, being in line with the longe-range aims of the Soreq-project, was chosen for this purpose.

The EURATOM Thesaurus is built on the principle of coordinate indexing for computerized systems. "It was devised with the view of compatability with the vocabularies of other organizations in order that it might serve as a basis for international co-operation"... - says the introduction to the second edition.

This guiding principle is clearly illustrated in the ratio between the number of 'authorized descriptors' (4.665) and the number of 'accepted keywords'. (11.030). The Thesaurus is intended to serve more as a bibliographical grouping devise than guide for constructing a subject index for particular collection. The omission of the use of role indicators is another sign of the higher generic level of its terms. For the purpose of general bibliographic orientation, the scope of the individual term is defined by its neighbouring concepts.

The design of the actual index

The best way for a closer explanation of the index is to give a general survey through typical examples.

1 Isotope separation / generic descriptor /

Isotope separation (Nuclear technology) 621.039.3
EURATOM
UDC

by chemical methods	621.039.32...
by physico-chemical methods	621.039.33...
by physical methods	621.039.34

(green)

For this type of entry a green card is used. The green colour and the equivalent form of the two sources quoted under the descriptor signify that the UDC and the EURATOM both recognize the concept on the same level.

The construction of the entry demonstrated the different functions of the two sources in the Soreq system: the descriptor 'Isotope separation' is fixed by the EURATOM Thesaurus, and UDC provides the subject posting through the hierarchic chain.

The subdivisions under the line do not in this case constitute descriptors, with separate cards in the index, but the next subdivision in the UDC number does have specific keyword value, as seen in Example 2.

2 Electromagnetic separation / specific keyword /

> Electromagnetic separation (Nuclear technology) 621.039.343
> EURATOM-see/or
> UDC

(white)

For this type of entry a white card is used, signifying a difference in the level ratings of the two sources. According to EURATOM the entry is given as an 'accepted keyword', but in the UDC it is a main-schedule concept.

No concept is entered into the vocabulary until monographic material has been collected on the subject. If a new subject rates publication in monographic form, this indicates that it has gained a measure of maturity. If in addition, the subject can be directly expressed through the UDC, the concept enters the vocabulary as a specific keyword. If however the subject has not yet reached a stage of development which qualifies it for inclusion in either the UDC or EURATOM as a fully fledged concept, despite the existence of monographs, it is dealt with as in Example 3.

3 Reactor stability / tentative keyword /

/Reactor stability 621.039.512/14
/kind of problems between
Reactor statics and kinetics
Hitchcock: Nuclear reactor stability

(white)

The card is white and no source is given, showing that the concept does not appear in the UDC or in the Thesaurus. The entry as a tentative keyword is an auxiliary classificational devise for internal use (red line in the left upper corner). This aspect is also indicated by the scope note.

4 Pressure vessels / generic descriptor - correlative function /
This item shows a different construction from the previous examples.

Pressure vessels (Reactor components) 621.039.536.2
EURATOM
UDC

In Light water moderated and cooled reactor 621.039.524.44.036.2
In Heavy water moderated and cooled reactor 621.039.524.45.036.2

(green)

The card is again green, and the two sources are quoted in an equivalent manner, indicating the equal level of the concept in the two authorities. But whereas in the example on Isotope Separation, EURATOM serves as a level-measuring tool, in the present case its correlative function is shown. This correlation is needed since, although the concept is a main-schedule term in the UDC, it can also be used as a special auxiliary, which would be a source of noise in retrieval. The concepts entered under the line form a series of independent specific keywords, and are headings for different aspect-groupings on the lower, specific level.

The collecting and organizing function of the Thesaurus is best manifested at the top level. These entries have only the authority mark of EURATOM on their headings.

5 Evaporation / collecting - top level descriptor /

a

Evaporation
EURATOM

Heat physics	536.423
Theoreticl chemistry	541.66

(orange)

The scope of this descriptor is very broad. In the EURATOM stock it is a term for 5,648 documents. Its broadness is expressed by the wide spread in the UDC notations under the line.

The cross references 'Heat physics', 'Theoretical chemistry', signify the various broader aspect-possibilities to which the concept can be connected. So according to the descriptor they are aspect-determinants. In the UDC language this means that in indexing, emphasis may be put on the classes (536 or 541) while the sections in the notation 536.423; 541.66 refer to Evaporation on the specific level, e.g.:

b

Evaporation (Heat physics) 536.423
UDC

(white)

In this example the former determinant appears in the heading of the entry as a supplement to the specific keyword. So in this case it gives the hierarchic posting of the term as a qualifier.

The above example shows how a term may be scaled down. The following one demonstrates how it can be raised to a higher level.

6 Nuclear evaporation - Evaporation models

In the Soreq collection there are several monographs on nuclear evaporation. In the Thesaurus the concept is a forbidden term (it is too narrow for the EURATOM subject policy). The idea can be expressed by a combination of two terms: 'Evaporation Models' and 'Nuclear Reactions'. The term 'Evaporation models' is itself only an accepted keyword and is more correctly expressed in the Thesaurus as 'Nuclear models' or 'Particle models'. Thus 'Nuclear evaporation' becomes 'Nuclear models' and 'Nuclear reactions'. The concept 'Nuclear evaporation' does not appear in the UDC schedules either. According to its intrinsic meaning it can be expressed in the UDC as the combination Nuclear models, 539.142 (from the class Nuclear structure, 539.14) in relation to Nuclear excitation, 539.17.012 (from the class Nuclear reactions, 539.17).

In the Soreq scheme the concept 'Nuclear evaporation' is expressed through three cards, based on the EURATOM Thesaurus. The first is a synonym card leading to the main card 'Evaporation models', which is a correlative card on the specific level. This card carries an aspect determinant 'Nuclear reactions'. The third card is an extra one for internal use (red diagonal line) which give the concept as a tentative keyword on the bottom level.

a

Nuclear evaporation =
Evaporation models (Nuclear reactions)
EURATOM

(white synonym card)

b

Evaporation models
EURATOM - see/or

Nuclear reactions: 539.142 : 539.17.012

(white card with specific keyword)

c

/Evaporation models (Nuclear reactions)
539.142 : 539.17.012

(white card with tentative keyword)

Now let us summarize what has been said.

Four levels of index terms are recognized: top, generic, specific and bottom. Each kind of term is distinguished by a different card colour. Each is also related to the kinds of term recognized by the EURATOM Thesaurus:

authorized terms, accepted terms, tentative keywords and forbidden terms.

Top level terms (collecting descriptors) are authorized terms from the Thesaurus that embrace more than one UDC class, such as Evaporation in the ex.5. Generic terms are authorized terms from the Thesaurus coinciding with a UDC class, such as Electromagnetic separation in ex.2. Lastly, bottom level terms are tentative keywords or forbidden terms in the Thesaurus, nevertheless needed in the index (ex.3. and 6.).

Excepting the synonym cross-references, each card in the index is divided into two parts. The upper half gives the index term corresponding UDC number if any, the authority or source of the term (EURATOM and/or UDC). The lower half gives references that may be transitions between hierarchic levels, or they can collect the components of correlative units.

The index consists of two files. One is organised according to the UDC, which provides a survey to the hierarchic context. The second file is alphabetical, based on the EURATOM Thesaurus. This one is the direct tool for retrieval.

The interplay of these two devices is manifested by the use of aspect determinants or qualifiers. The headings of the UDC classes serve this purpose. If such heading appears in parentheses beside a generic descriptor or specific keyword it serves as a qualifier and so determinates the hierarchic posting of the term. The same expression appearing under top level collecting descriptor - as it was the case with Heat physics and Theoretical Chemistry in the Evaporation - example - signifies the various subject fields to which the concept can be related. In this case the qualifier becomes aspect-determinant.

Substantially this is the simple device through which - in combining two conceptually different systems - a virtually new one was created.

The described correlative index is not a speculative proposition. It is a daily used instrument of information-supply at the service of the Soreq Nuclear Centre's scientists.

REFERENCES

1. BREE, R. Erfassung und Verbreitung kerntechnischer Information. Atomwirtschafte-Atomtechnik, vol. xii, no. 7, July 1967, p. 365-9.
2. EURATOM-Thesaurus. Indexing terms used within EURATOM's Nuclear Documentation System. 2nd ed. 1966, EUR 500e.
3. INTERNATIONAL ATOMIC ENERGY AGENCY. Working Group on International Nuclear Information System. Vienna. PL-233/1 1966: PL-233/18/Rev. 1.
4. KORNER, HORST. Short and flexible generic codes for mechanized information retrieval In: International Conference (on) Classification Research. 2nd Elsinore, 14-18 Sept. 1964. Proceedings, ed. PAULINE ATHERTON. Copenhagen, Munksgaard, 1965, p. 268-91.
5. MAROSI, A: EURATOM Thesaurus and UDC. Journal of Documentation, vol 25, No 3. Sept. 1969. 197-213 p.
6. MEYER-UHLENRIED, K. H. Classification and automatic documentation. (EURATOM research. In Int. Conf. (on) Classification Research. 2nd Elsinore, 14-18 Sept. 1964. Proceedings, ed. by Pauline Atherton. Copenhagen, Munksgaard, 1965, p. 326-40.
7. ROLLING, LOLL N. Computer aided information service. Journal of Documentation, vol. 22, June 1966, p. 93-115.
8. ROLLING, LOLL N. The role of graphic display of concept relationships in indexing and retrieval vocabularies. In Int. Conf. (on) Classification Research. 2nd Elsinore, 14-18 Sept. 1964. Proceedings, ed. by Pauline Atherton. Copenhagen, Munksgaard, 1965, p. 295-325.
9. RUSSELL, MARTIN and FREEMAN, ROBERT R. Computer aided indexing of a scientific abstracts journal by the UDC with UNIDEK. A case study. American Institute of Physics, 1967 (AIP/UDC-4).
10. Universal Decimal Classification. Special Subject Edition for Nuclear science and technology. The Hague, FID, 1964.

1. **Isotope Separation.** /generic descriptor/

Isotope Separation (nucleat technology)	621.039.3
EURATOM	
UDC	
by chemical methods	621.039.32...
by physico-chemical methods	621.039.32...
by physical methods	621.039.34...

green

2. **Electromagnetic Separation** /specific keyword/

Electromagnetic Separation (nuclear technology)	621.039.343
EURATOM - see/or	
UDC	

white

3. **Reactor Stability** /tentative keyword/

Reactor Stability	621.039.512/14
/kind of problems between Reactor Statics and Kinetics.	
Hitchcock: Nuclear Reactor Stability	

white

4. Pressure Vessels

This item shows a different construction from the previous examples. /generic descriptor-correlative function/

Pressure Vessels (Reactor Components)	621.039.536.2
EURATOM	
UDC	
In Light Water Moderated and Cooled Reactor	621.039.524.44.036.2
In Heavy Water Moderated and Cooled Reactor	621.039.524.46.036.2

green

5. Evaporation /collecting - top level descriptor/

a)

Evaporation	
EURATOM	
Heat Physics	536.423
Theoretical Chemistry	541.66

orange

b/ /specific keyword

Evaporation (Heat Physics)	536.423
UDC	

(White card)

6/ a)

Nuclear Evaporation =
Evaporation Models (Nuclear Reactions)
EURATOM

(White synonym card)

b)

Evaporation Models
EURATOM - see/or
Nuclear Reactions: 539.142:539.17.012

(White card with specific keyword)

There is no UDC notation above the line, indicating that this is a collector card.

c)

Evaporation Models (Nuclear Reactions)
539.142:539.17.012

(White card with tentative keyword)

TABLE I: STRUCTURE OF THE SOMEO INDEX

LEVEL	TYPE OF TERM	AUTHORITY	COLOUR OF CARD	DEVICES FOR INCREASING SPECIFITY OF TERMS	FUNCTION OF SUBDIVISIONS FROM POINT OF VIEW OF THE HEADING, ACCORDING TO ROLE OF EURATOM/UDC IN THE SYSTEM	EXAMPLES
Top	Collecting Descriptor*	EURATOM	Orange	aspect determinants or qualifiers according to the class or section headings of the UDC	Devices for transition to the next stage. For EURATOM headings - aspect determinants For UDC subdivisions - qualifiers	Ex. 5
Generic	Descriptor	EURATOM/UDC	Green	aspect determinants and qualifiers (overlapping)	a. Concept unit Heading (EURATOM) - diameter of concept Sub.div. (UDC) - hierarchic posting Structural function without keyword value	Ex. 1)
					b. Correlative concept Heading (EURATOM) - hierarchic posting Sub.div. (UDC) - collecter of elements in independent keywords for narrower level	Ex. 4)
Specific	Independent Keyword**	UDC and/or EURATOM accepted term	White	qualifiers	a. Concept unit Heading (UDC and/or EURATOM accepted term) - diameter of concept Sub.div. - tentative keyword for the bottom level	Ex. 2
					b. Correlative concept Heading (UDC and/or EURATOM accepted term) - determination of the specific level Sub.div. - independent keywords on the same level	Ex.4
Bottom	Tentative Keyword	———	White with lines	Scope notes, if required		Ex.3 Ex. 6

*Descriptor: EURATOM "authorised term"

**Keyword : UDC term and/or EURATOM "accepted term"

OPTIMIZING THE STRUCTURE OF CLASSIFICATION AND INDEXING SYSTEMS

Ladislav KOFNOVEC, Dušan SIMANDL
Centre of Scientific, Technical and Economic Information,
Prague, Czechoslovakia

In the field of information science and also in the information and documentation practice the last decades are characterized by the general trend to utilize widely the modern technical equipment, above all the computers. The initial fascination by the modern technical means was soon replaced by a sober evaluation of advantages and disadvantages of the broad application of modern technical means in comparison with the conventional methods. In the present stage of a more realistic evaluation of modern methods it is shown that a sufficiently effective indexing and retrieval language is a very important condition for reaching a desirable efficiency of an information retrieval system (whether automatized or conventional).

Even though the majority of modern discipline-oriented computer-based information systems use the descriptor languages (mostly based on the natural language), we find here also certain number of applications of other than descriptor languages, e.g. the UDC. It was particulary difficult to substitute efficiently the descriptor systems in connection with the organization of polythematic (multidisciplinary) files.

One of outstanding and recurrent features of today's efforts in the field of information work are ideas on building-up integrated networks of information centres and libraries, whether on the national, regional, or even international level. In conection with this it is clear that the individual components of such a system have to be interlinked by a common switching language which probably should have a much higher level of generality than the specialized indexing and retrieval languages of the individual information subsystems. With regard to the fact that up to now no attempt to develop a universal polythematic descriptor language has met with success it can be said that, probably a hierarchical classification system will be most suitable for the "linking" role.

Therefore it is clear at present that the UDC could become in the near future a "roof" system enabling the linking of individual discipline-oriented subsystems within the framework of national, regional or international information macro-systems on the one side or of micro-systems specialized in certain types of documents etc. on the other side. Polythematic library and information collections will probably remain an important field for the use of the UDC, but the UDC will certainly have an important role also in specialized information centres and special libraries, notwithstanding the present trend to computerized discipline-oriented information systems which use to a great extent descriptor systems as retrieval languages.

With regard to this perspective of the UDC we are convineed that it is necessary to concentrate the greatest effort on its further development and systematic improvement and that it is also necessary to study its features and characteristics. One of the most important characteristics of every retrieval language is its structure and the thorough knowledge of this structure can help us to make a better assessment of potential applicability of such a language.

Unfortunately we find very few information in the today's literature on the investigation of the structure of the world scientific and technical literature which constitutes in fact an universe that must be organized by means of a classification system with the objective of a future retrieval of the required information or documents. In our view it is possible to ascertain the suitibility of the classification system's structure only so that we shall compare it to the most objectively ascertained structure of the world scientific and technical literature. Naturally this is even more important in the case of a universal classification system.

Of course, such a quantitative approach has its shortcomings and limitations. It enables, however, to evaluate a system in broader connections, to make the necessary comparisons, and gives objective data for planning revisions and corections of the whole system or of certain parts of it.

With the objective of a better appreciation of the UDC's role we investigated the structure of the world scientific and technical literature and the structure of the UDC on the level of the full edition and of some other Czechoslovak editions. We should like to inform on some results of the investigation on the UDC's structure, because the information on the world scientific and technical literature's structure will be published elsewhere (1). The following data illustrate the results of the investigation of the UDC full edition according to "Extensions and Corrections to the UDC" Series 7, No.2.

The fundamental ascertained break-down of the UDC numbers - on the class level - in the full edition is as follows:

UDC class	UDC numbers					%
	Common auxiliaries	Main numbers	Special auxiliaries	Relation (colon combination)	Total	
Common aux.	8 083		132	5	8 220	6,6
0		971	520	58	1 549	1,2
1		1 361	348	23	1 732	1,4
2		1 982	282	25	2 289	1,8
3		10 543	2 429	425	13 397	10,7
5		20 775	2 786	288	23 849	19,1
6		41 833	24 142	1 299	67 274	53,9
7		2 961	1 433	88	4 482	3,6
8		699	223	1	923	0,7
9		602	627	1	1 230	1,0
	8 083	81 727	32 922	2 213	24 945	100,0

Common and special auxiliaries (in the latter case actually compound numbers with special auxiliaries) were counted only once. Special auxiliaries with identical notation but added to another main number were counted more than once only when their texts were different. Further, 667 common auxiliaries were found which were used in the main table in connection with main numbers and 4 509 special auxiliaries which were used more than once in the same or very similar sense. These repeated UDC numbers represent 4% of the total. Exceptionally they exceed 5% and only in the case of 621.7/.9 they exceed 10%. Therefore they do not substantially influence the results.

The structure of the UDC according to the types of UDC numbers is given below:

Type of UDC number	Number of UDC numbers			%		
	Basic use	Repeated use	Total	Basic use	Repeated use	Total
Main numbers +/	81 727	-	81 727	65,4	-	62,8
Common aux.	8 083	667	8 750	6,5	12,9	6,7
-1/-9	3 621	386	4 006	2,9	7,4	3,0
.01/.09	24 367	1 931	26 298	19,5	37,4	20,2
-0, -00	4 334	2 191	6 525	3,5	42,3	5,0
,	407	2	409	0,3	-	0,4
...1/...9	193	-	193	0,1	-	0,2
Relation	2 213	-	2 213	1,8	-	1,7
	124 945	5 177	130 122	100,0	100,0	100,0

The "repeated use" occurs only in the case of common and special auxiliaries in the sense previously mentioned.

The appended Table 1 is a survey of such a detailed distribution which is necessary for the quantitative comparison of the UDC structure with the structure of the world scientific and technical literature, i.e. to the degree of detail we were able in the first stage of our investigation to ascertain from corresponding data on the importance of the individual subject fields according to the scope of published significant documents in the fields of science and technology. Apart from this the table also shows the numerical breakdown of individual types of UDC numbers in each of the mentioned sections. The repeatedly used common and special auxiliaries were not included in this table.

Further considerations of the most correct evaluation of the totals of UDC numbers which should form the basis of comparison brought some other justified requirements and recommendations the realization of which was not possible, sin-

ce up to now those data which practically could only be a product of another independent research are not available. But at least we shall try to describe how to proceed in the possible application of these criteria:

It would certainly be correct to assess the weight of the individual special auxiliaries differently, probably higher than of the main numbers. We assign them a higher importance on the basis of experience and also by intuition, since one can anticipate their more extensive use in connection with other main UDC numbers along with those main numbers where (in the full edition) the special auxiliaries are explicitly quoted. The coefficient K_s which would determine the average weight of the special auxiliary (supposing the weight of the main number is 1) could have - according to our guess- perhaps the value 3.

Similarly it would be useful to evaluate the average weight of the common auxiliary (possibly for their individual types separately) by the coefficient K_c. Without the necessary data we do not dare to guess the value of this coefficient. Moreover it seems that this value would differ from one section to the other.

If we want to show how many cards are in average per 1 number - supposing an average of a more or less current use of more than one number for indexing of every document- it would be necessary to introduce another coefficient K_d, by which either the number of yearly published important documents would be multiplied or the number of the respective UDC numbers would be multiplied by its reciprocal value.

The above stated considerations are complicated by the fact that the situation is changing from one section to the other according to the character of the subject field and the structure of its arrangement in the classification scheme and that it would be necessary to use various values of the proposed coefficients for various parts of the UDC.

We have reasons to assume that after a wider practical application of the SADF[x)] programs in the determination of structures of a sufficient number of actual information files in Czechoslovakia it would be possible to use the obtained data for the more objective description of the structure of the UDC.

In Table 2 we present - according to UDC classes or sections- the actually ascertained or guessed numbers of yerly published significant documents which must be considered for abstracting. Column 3 contains the data on UDC numbers in the full edition. The detailed report also shows i) calculated average numbers of documents per 1 UDC number, ii) the specifity of classification which is expressed by the reciprocal value of the average number of documents per 1 UDC number multiplied by 100.

The overall comparison shows that in the case of the full edition there were nearly 11 significant documents per 1 UDC number.

Comparisons and conclusions should be usable for the considerations on priority fields where revisions should be started as well as for precising the basic quantitative principles for the conception of the possible future "Scheme of UDC numbers"[xx)], although in the latter case it would be necessary to consider some other important corrective factors.

---

x) Statistical Analysis of Documentation Files. Further information on these programs can be found in papers published in "Metodika a technika informaci" (2,3).

xx) This is a polythematic UDC edition with differentiated totals of UDC numbers in individual sections according to their importance for individual Czechoslovak information centres and libraries.

As a basis for further considerations we show a table which will demonstrate where the main weight is concentrated:

UDC class or section	UDC numbers	
	Full edition	
	Total	%
6	71 269	54,76
62	19 695	15,18
621.3	4 415	3,39
621.7/.9	3 354	2,58
66	12 557	9,65
663/669	8 272	6,36
67/68	13 811	10,61
5	24 286	18,67
3	13 681	10,51

This table shows that in classes 6, 5 and in the full edition over 83 % of all UDC numbers are concentrated (with the exception of common auxiliaries). With regard to the weight of these classes and owing to the fact that our interest is first of all concentrated on the field of scientific, technical and economic literature, we shall focus our attention to these fields.

But let us compare the above-stated thematic fields with the respective weight of the scientific and technical literature:

UDC class or section	Estimated numbers of yearly published significant documents (in thous.)
6	650
62	240
621.3	65
621.7/.9	60
66	135
67/68	45
5	400
3	90
Total	1 140

We obtain rather a similar picture as in the investigation of the structure of UDC numbers. The thematic fields represent 1 140 thousand significant documents from 1 280 thousand significant documents that are published yearly, i.e. 89 %, which is more than the share of numbers in the full edition.

If we take as a basic bracket for the full edition 6 to 19 documents per 1 number, where - as we found out earlier - more than one half of the stated fields is concentrated, we can see that in the classes under consideration (i.e. 3, 5 and 6) the upper limit of the bracket was exceeded in the following thematic fields:

UDC section	Documents per one UDC number	Exceeding of fixed limit
30	25,21	6,21
31	46,69	27,69
32	24,68	5,68
50x)	131,57	112,57
53	25,32	6,32
57	63,69	44,69
621.22	42,85	23,85
621.7/.9	20,59	1,59
69	23,34	4,34

x) But see the instruction in the full edition about the rather extensive use of colon combinations

Although the other classes - except 3, 5 and 6 - are less or only partly interesting for us, it is necessary to say that the fixed limit was also exceeded in class 8xx) (32,5 %) and class 9 (24,39 %).

The data for the full edition point to priority revisions(i.e. further subdivision is necessary) above all in those sections where the number of documents per 1 UDC number exceeds considerably the fixed limit. This applies above all to sections 31, 57 and 621.22, in the second place to sections 30, 32, 53 and 69; the exceeding of the limit in the case of 621.7/.9 is negligible.

---

xx) Though our interpretation of the number of yearly published documents in this class is strongly questionable.

The described results ensue from the quantative approach chosen - with all its shortcomings and limitations. There was no attempt made to investigate the usefulness of existing subdivision of individual subjects in relevant sections. Thus with the approach chosen the otherwise evident need for elaborating the fields of cybernetics, automation (and information science, too), solid-phase physics, certain sections of biology, geology, chemistry, electronics etc. was not established. It is probable, however, that these facts could be established also with the quantitative approach, but on the micro- and not macro-level as it was done in the investigation described above.

The overcrowding of class 6, which was pointed out e.g. by Lloyd (4) in his estimate of the full edition's structure, results - among others - in excessively long UDC numbers in certain parts class 6 (e.g. in 621.3). At the same time the amount of information in the field of technology is increasing considerably. This is illustrated e.g. by the following data on the growth of the number of abstracts in certain abstracting journals:

Journal	Number of abstracts		Growth %
	1957	1967	
International Aerospace Abstracts	6 770	33 116	389
Nuclear Science Abstracts	14 000	50 000	251
Review of Metal Literature (now Metals Abstracts)	8 219	23 800	190
Chemical Abstracts	101 028	240 000	138

The long discussion about improving the situation in lass 6 with regard to the vacancy in class 4 is still continuing. According to a recent proposal of the Classification Department of FID (6) enough space in class 6 could be obtained for the rapidly developing technical disciplines by transferring certain parts of class 6 to class 4 or 3.

According to this alternative the following changes in class 6 were proposed:

Action suggested		Future possibilities	
6	Medicine, husbandry & technology (Applied science)	6	Technology & industrial production
x61	(Medical science) → 46		- free for Materials techn. or other
62	Engineering & technology (generally)	62/64	Energy technologies
		62	Mech. engng, machines & structures
x63	(Agriculture, etc.) → 47	+63	Electrical engineering. Electronics
x64	(Household management) → 38	+64	Nuclear & other energy technologies
x65	(Business management) → 38 exc. 656	!65	Transport engineering & services
66/68	Chemical tech. & manufacturing indust.	66/68	Chemical & manufacturing indust.
69/72	Building, physical planning & archit.	69/72	Building, civil engng, physical planning & architecture

Conclusions

In the future it is necessary to concentrate an extraordinary effort on the investigation of the structure of the world scientific and technical literature and on its comparison with the existing structure of the UDC. Establishing both structures will constitute an important basis for an objective outline of the further development of the UDC and its individual sections and for establishing major priorities in the revision works.

Establishing the structure of the world scientific and technical literature is also necessary for the further development of information science as a discipline, since it describes the existing structure of scientific, technical and economic information which is the basic object of study of the information science.

It will be possible to use the results of the comparison of both structures for determining the desirable optimum relations of the scope of individual sections in the abridged, medium and full UDC editions, as well as in special subject editions (supposing a suitable level in the investigation of the structure).

With regard to an extraordinary importance of this work the proposed investigation in the above described sense should be organized directly by the FID in cooperation with the national FID members through the relevant classification and study committees, e.g. FID/CCC and FID/RI.

Literature

1. SIMANDL, D. - KOFNOVEC, L.: Sovremennaja struktura mirovoj naučnoj i techničeskoj literatury. (Present structure of the world scientific and technical literature.) Nauč.-techn. Inform. (in print)

2. MOJŽÍŠEK, J. - MAIXNER, V.: Statistická analýza dokumentačních fondů. (Statistical analysis of documentation files.) Met. Techn. Inform., 12, 1970, No. 12, p. 32-44

3. MOJŽÍŠEK, J.: Ditto. 2nd part. Českoslov. Inform., 13, 1971, No. 1, p. 26-39

4. LLOYD, G.A.: Naturwissenschaft und Technik in der künftigen Revision der DK. DK-Mitt., 9, 1964, No. 1, p. 4-9

5. Scientific and technical communication. Washington, Nat. Acad. of Sciences 1969.

6. Document FID No.C 70-17. The Hague, FID, April 28, 1970

Table 1

UDC class or section	UDC numbers							
	Main numbers +/	Special auxiliaries				...1/ /...9	Relation (colon combinations)	Total
		-1/-9	.01/.09	-0, -00	'			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Common auxiliaries	8 083	58	68	6			5	8 220
0	971		516	4			58	1 549
2	1 982	163	87	32			25	2 289
3	10 543	80	2 296	53			425	13 397
30	165	37	21	12			3	238
31	135	21	3	2				161
32	349		34	5			17	403
33	1 882		418	4			168	2 472
34	3 771		257				43	4 071
35	1 950	7	446	6			19	2 428
36	798		591	17			119	1 525
37	834	15	522	7			47	1 425
38	357						4	361
39	302		4				5	311
5	20 775	327	2 211	7	226	15	288	23 849
50	13						25	38
51	2 190	6	40				2	2 238
52	956	78	328				36	1 398
53	3 155	121	575			15	43	3 949
54	5 499	107	442	7	226		44	6 325
55	2 432		340				112	2 884
56	1 237		23				4	1 264
57	898	10	346				2	1 256
58	1 877		58				1	1 936
59	2 478	5	59				19	2 561

Table 1 (cont.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
6	41 833	2 767	16 794	4 231	181	169	1 299	67 274
60	9							9
61	6 043	7	353	4 054			121	10 578
62	12 129	2 036	4 225	1		17	496	18 904
62/62-	1	1 792						1 793
620	608	2	24				3	637
621	6 628	99	2 330			17	210	9 284
621.0	281		28			17	3	329
621.1	469	3	28				2	502
621.2	39	20	11					70
621.3	2 930		1 212				125	4 267
621.4	140	63	327				1	531
621.5	214	1	36				2	253
621.6	378	1	39					418
621.7/.9	2 177	11	649				77	2 914
622	1 230	49	240				92	1 611
623	676	5	135	1			28	845
624	683		218				37	938
625	678	2	131				5	816
626	274	21	14				8	317
627	366	4	78				9	457
628	673	1	68				103	846
629	312	61	986				1	1 360
63	5 473	2	813	9	6		86	6 389
64	1 044	5	77				49	1 175
65	2 276	14	1 177	164			198	3 829

Table 1(cont.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
66	8 162	562	2 974	3	19	150	261	12 131
67	1 946	76	5 284		156		75	7 536
68	4 153	56	1 647			1	9	5 866
69	598	9	244			1	5	857
7	2 961	49	1 374	1		9	88	4 482
8	699	168	55				1	923
9	602	2	625				1	1 230
Common auxiliaries +0/9	89 810	3 621	24 367	4 334	407	193	2 213	124 945

Table 2

UDC class or section	Number of documents per year (in thousands)	UDC numbers in the full edition (in thousands)
(1)	(2)	(3)
0	15	1,549
1	25	1,732
2	10	2,289
3	90	13,397
30	6	0,238
31	8	0,161
32	10	0,405
33	25	2,472
34	10	4,071
35	8	2,428
36	5	1,525
37	8	1,425
38	6	0,361
39	4	0,311
5	400	23,849
50	5	0,038
51	23	2,238
52	15	1,398
53	100	3,949
54	70	6,325
55	50	2,884
56	2	1,264
57	80	1,256
58	10	1,936
59	45	2,561
6	650	67,274
61	100	10,317
62	240	18,904
62 (introduction)	7	1,793

Table 2 (cont.)

(1)	(2)	(3)
620	7	0,637
621	150	9,284
621(introduction)	5	0,329
.1	4	0,502
.22	3	0,070
.3	65	4,267
.4	8	0,531
.5	2,5	0,253
.6	2,5	0,418
.7/.9	60	2,914
622	25	1,611
623	10	0,845
624	10	0,938
625	8	0,816
626/627	3	0,774
628	5	0,846
629	24	1,360
63	70	6,389
64	10	1,175
65	30	3,829
66	135	12,131
67	25	7,536
68	20	5,866
69	20	0,857
7	30	4,482
8	30[x/]	0,923
9	30	1,230
0/9	1 280	116,725

[x/] Without fiction

UDC-INSTRUMENT FOR CONNECTING SCIENTIFIC INFORMATION, EDUCATION AND PRACTICE

Milan MOLE

The Yugoslav Building Centre,Beograd,Yugoslavia

(1) Dissemination, storage or input(in the broadest sense of the word) and retrieval or output of scientific information(as a basis for the acquisition of knowledge) on particular facts and phenomena requires subject indexing of each piece of information on the input-side and of the inquiry on the output-side.

(2) If knowledge is produced by the science, more precisely by several scientific disciplines, each of them with its own aspects of investigating natural and social facts and phenomena on things,products, ideas etc., and with its own classification scheme - then these classification schemes appear the only logic schemes for the classification of both scientific information and inquiries in each scientific discipline. Each such scheme is in fact a faceted and/or hierarchical classification scheme, and the UDC is in fact (or it should be) the sum of such classification schemes for all scientific disciplines, i.e. for the whole field of knowledge.

(3) UDC notation rules for the systematic main tables render possible the identification of each scientific discipline and separation of their classification schemes. (It is already pointed out that for the "internal" use of the UDC in one field, the first figures - the notation for the field itself - could be left out). The always necessary alphabetical indices to the UDC show what subjects are (already) treated by several scientific disciplines or their parts.

(4) Some experts consider that an enumerative or hierarchical or faceted or combined list of all existing and new scientific disciplines or such an extract from the UDC, on the one side, and the special thesauruses for each scientific discipline, on the other side, would be the necessary instruments that would make the UNISIST feasible.

(5) One must bear in mind the fact that scientific information should reach not only the scientists of their own disciplines but, first of all, it should reach men of practice who will apply it. It is also essential that scientific information should reach the scientists of other branches, especially in multidisciplinary teams in research and social missions.

(6) Men of practice belong to different professions, each of them using (needing) knowledge from a great number of sciences. This begins already with the training when knowledge of different intensity from these different sciences is assimilated, which lasts then during one's whole professional life. Is this not a very strong argument to give each educational subject, i.e. each science or part of it, always the same classification mark? Under this mark the practitioner should receive the introductory knowledge in school, the references on new knowledge in the information bulletins for his profession or his branch of activity, respectively, through the SDI.

(7) It is more and more evident that the scientific information, especially in its condensed form, accompanying or preceding the full scientific information, ought to be repacked for different user(need) groups. Each package will always consist of scientific information from several sciences. It is essential to have the continuity of knowledge in his brain, which should be organized and in fact is organized through the education. The best organizational scheme for the knowledge in one`s brain is exactly the UDC, with its enormous feasibility of adapting itself to a countless number of "profiles", by shortening the classification marks in accordance with individual needs. Thus, abridged and full special editions of the UDC, printed out automatically from the UDC master file, would be the excellent organizational schemes of knowledge for each profession, or for each whole branch of activity (where more professions are acting).

(8) This UDC master file, the classification scheme for all sciences, is the only effective common denominator for the knowledge in the brains of all professions and branches of activity, and the only instrument for the inter-connection of sciences, and for the connection of research, education and practice.

(9) Each branch of activity and each profession needs several kinds of information, among others scientific information, if possible, specially fassioned(adapted or "addressed") to each. Sometimes it is useful to point out that scientific information is adapted specially for one branch or one profession. Classification schemes for these categories does exist (UN and ILO), and they should be used adequately.

(lo) It is necessary to introduce in the UDC the auxiliary numbers to denote the intensity of information, e.g.:
- first announcement by the author(s) - "notice scientifique"
- exhaustive, full presentation by the author(s) - "mémoire scientifique"
- revised full information
- adapted information, addressed to : (profession, branch of activity)
- introductory information
- general information : for interdisciplinary connection
  popularization

This is only a first approach. - These numbers could be placed,perhaps, together with auxiliaries for the form, or they could be a special facet in the series of -0 numbers.

(11) From all the reasons, drafted as above, UNESCO should support the development of the UDC, as the only international science indexing language.
The fast up-dating and the purification of the UDC-main tables, the editing of abridged and full special editions based on the up-dated UDC master files, in several languages, will provide maybe the only chance for the implementation of the UNISIST-idea within the field of application of science all over the world.

(12) Finally, it should be pointed out that no other indexing language is able to play the role of the necessary common denominator for the classification of knowledge both on the producer-side (science) and on the user-side (education and practice), on the national level as well as internationally, thanks to the ingenious use of the decimal system of numbering.

A THEORETICAL FOUNDATION FOR UDC: ITS NEED AND FORMULATION

A.NEELAMEGHAN
Documentation Research and Training Centre,Bangalore 3, India

0 INTRODUCTION

01 Terminology

011 Indexing language

For the purpose of this paper, the term 'Indexing Language' is taken to denote a scheme for classification of subjects embodied in documents and an associated system of subject heading.

012 Classification of Subject

The classification of a subject consists essentially of

1 Recognising the component ideas in the subject;

2 Determining the appropriate degree of interrelation among the component ideas; and

3 Arranging the component ideas in a linear sequence according to their mutual filiation as determined by the degree of interrelation among them and preferred by a majority of the specialists in the subject.

013 Other terms

Some of the specialised terms used in this paper are defined in Appendix 1. These are taken from the Prolegomena to library classification, Ed 3 (1967).

1 UBIQUITOUS ROLE OF LIBRARY CLASSIFICATION

11 Arrangement of Books by Subject

Nearly a century ago, the genius of Melvil Dewey brought into almost univesal use the concept of arranging books in a library according to the subjects embodied in them. For this purpose, he designed a classificatory language of ordinal numbers, which is a time-honoured tool for any arrangement and maintenance of it. This is the Decimal Classification. It proved helpful. For, the arrangement of books by subject satisfied in a greater measure than by other methods, the approach of the majority of readers to the use of documents.

12 Classified Arrangement of Entries in a Catalogue

At the turn of the present century, Paul Otlet and Henri La Fontaine pioneered the idea of arranging the main entries in a vast catalogue of documents -- that is, a universal bibliography -- according to a scheme for classification. They sensed that a micro subject, such as that embodied in a section of a book or an article in a periodical, included in the bibliography, consisted of several units of different kinds of minute ideas in a way different from the subject of a whole book. To facilitate a close classification of such a subject, Dewey's Decimal Classification was fitted with the Facet feature -- Space Facet and Time Facet each with its own indicator digit, and the colon being used for all other kinds of facets and phases of subjects. Such a classified arrangement of entries has become a helpful tool in the search and selection of micro documents, while the micro documents themselves may be physically scattered in different host documents. The Decimal Classification fitted in this way for use in the universal bibliography was named Universal Decimal Classification. Alternatively, the epithet 'universal' might, perhaps, have been used because the new scheme for classification was ostensibly designed for the classification of micro subjects also.

The two kinds of use of library classification mentioned above are now well established and widely practised. However, we may note here that it is not unusual to find classificationists treating the work of designing of schemes for classification for macro documents and those for micro documents as if they should be independent and mutually exclusive.

13 Formulation of Subject Heading

Till about 1935, classifying the subject of a document and the preparation of subject headings for it have been largely done as if they were independent and mutually exclusive. Perhaps, even today this is the way it is practised in some large libraries. With the development of the technique of Chain Indexing by S R Ranganathan in 1934 (1), it became evident that a subject heading can be formulated by a translation into a meaningful representation of, the class numberin the verbal plane. Further, the use of Chain Indexing on a large scale -- for example, in university library catalogues, the British national bibliography, and the British technology index -- established its value as helpful in structuring a subject in the verbal plane and in formulating subject headings in a consistent, systematic, and economical way. In 1964, Ranganathan showed that the steps in classifying a subject according to the systematic procedure based on postulates and principles and those for formulating subject headings for it are identical, and therefore, could be guided by the same set of guiding principles for facet analysis (2). Essentially, the steps involve the recognition of all the fundamental constituent ideas in the subject, determining the different degrees of interrelation among them, and arranging the constituent ideas in a sequence helpful to the majority of specialists in the subject concerned. It has been further pointed out that the procedure based on the postulates and principles of classification need not be carried to the ultimate stage of constructing the class number and the result obtained in the penultimate stage could be conveniently used to formulate subject headings. Thus, the intimate symbiosis between classification and formulation of subject heading is firmly established (3).

14 Facet Analysis of Reader's Query

The subject of a reader's query can be facet analysed and structured in the same way and using the same set of guiding principles as those used in the facet analysis and structuring of the subject of documents included in a document finding system (4). In fact, this helps in matching the subject of the reader's query with those of documents to any pre-determined degree of coextensiveness. In document finding this helps in minimising noise and leakage. The more we refine the postulates and develop new guiding principles whenever necessary the more efficient the system becomes.

15 Reader-Librarian-Document Finding System-Trilogue

151 Deficiency in Query Statement and Query Analysis

The statement of a query by a reader may not represent co-extensively and precisely the subject of his interest at the moment. He may recall only a few of the component ideas. If he is working in the wave-front of his field of specialisation, his findings will often be so new that they cannot be fully expressed in terms of the existing terminology of his subject. Therefore, he is obliged to express his requirement in a non-coextensive terminology. These factors make the librarian's grasp of all the facets of the specifc subject of the query, on the basis of the statement of it by the reader, inadequate (5). These inadequacies in query statement and query analysis can lead to noise and leakage in document findings

152 Need for a Trilogue

An intimate trilogue between the reader, the librarian, and the document finding system is helpful in the formulation of the subject of the query as closely as possible. A carefully planned trilogue can help

1 The reader to

11 Recall, as many as possible, the component ideas of the subject of his interest at the moment.

[Annotation.- If the number and variety of ideas making up the subject are beyond the "span of immediate recall" it would be comparatively more difficult for the reader to remember them all (6). The time lapse between the stimulus to know and the expression of the need and the psychological characteristics of the reader also influence the situation.]

12 Conceptualise the component ideas in a way to secure coextensiveness between the verbal plane and the ideas denoted by it; and

2 The librarian to

21 Understand precisely and unerringly the reader's subject interest at the moment;

22 Supply any missing facet in the reader's query such that the document selection becomes pinpointed,

23 Arrange the component ideas of the subject derived during the trilogue in a sequence to represent nearly coextensively the reader's subject interest at the moment,

24 Transform the facet-analysed query into the language of the document finding system so as to facilitate the search and selection of entries for the appropriate documents, and

25 Modify the search strategy, as necessary, on the basis of

the continuous feed-back arising at the different stages of the reader-librarian-document finding system-trilogue.

153 Stages in the Specification of the Query

The precise formulation of the reader's subject interest at the moment may involve the following stages:

1 Rough enunciation, by the reader, of the subject about which he needs information;

2 Step by step facet analysis of the subject by the librarian;

3 If found helpful, display parts of the schedules of a well structured scheme for classification to help the reader to select the appropriate component ideas and to provide cue to help him recall any other component idea;

4 Consulting the alphabetical subject index to the catalogue or documentation list for a closer enunciation of the subject and to pick out the class number for it;

5 Landing the reader in the classified part of the catalogue or documentation list in the pack of entries bearing the class number picked up from the alphabetical part and covering or coming close to the subject of his interest at the moment;

6 Reader browsing among the classified entries provided with adequate number of feature headings and which display the subjects in a helpful sequence according to their mutual filiation; and

7 Narrowing down the search range to as small a pack of entries as possible on the basis of the feed-back while browsing.

This trilogue should help in a pinpointed selection of documents of interest to the reader at the moment, indicate the uncultivated areas in the subject, and help the reader in recalling any other subject or idea associated with his specific subject of interest at the moment. The value of an efficient scheme for classification in this trilogue has been dealt with in other papers (7).

16 Record of Reader's Subject Interest

From the discussion in the preceding sections, it becomes obvious that the same facet analysis technique and the set of guiding principles for it can be used with advantage in the precise formulation and co-extensive representation of a reader's subject interest in his profile record for use in Selective Dissemination of Information (8).

17 Discovery of Areas for Research

A well structured scheme for the classification of subjects going with a Basic Subject can be used to discover the combination of ideas on which work has not been done.

18 Synergistic Effect

Thus, classification -- that is, the analysis of a subject into the component ideas, recognising the degree of inter-relation among them, arranging the component ideas in a sequence helpful to the majority of readers, and the coextensive representation of the subject either in the form of a subject heading or as a class number -- is incident, in one way or the other, in several stages in the design, development, and use of a document finding system. Therefore, the efficiency of the document finding system depends, in a large measure, on the efficiency of the scheme for classification used. Any improvement in the efficiency of the scheme for classification would effect improvements in the efficiency of the document finding system as a whole several fold -- that is, there will be a synergistic effect. Further, a "living" dynamic scheme for classification will keep up with the pressures of the universe of subjects and of the psychology of readers. Therefore, the system for document finding using such a scheme for classification is equipped with a built-in capacity to meet the challenges of the developments in the universe of subjects and psychology of readers in an efficient way. In the design, development, and use of a document finding system this ubiquitous and basic role of classification should be kept in view.

2 ATTRIBUTES OF THE UNIVERSE OF SUBJECTS

The attributes of the universe of subjects which affect the design and development of a scheme for classification and the mental set of the user -- be it the specialist reader or be it the librarian -- of a document finding system relate mainly to its structure and development. These attributes include:

1 Patterns of combination of constituent ideas in subjects -- that is, modes of formation of subject; and

2 Dynamic growth resulting in the universe of subjects tending to infinity and a continuum.

These attributes have been discussed in some detail in other papers (9). Therefore, some of the salient points alone are mentioned in the succeeding sections.

21 Modes of Formation of Subject

A subject is essentially a systematised body of ideas embodied in documents to facilitate communication. These ideas may be of different varieties. The pattern of combination of ideas making up a subject may be of different varieties. Each such pattern of combination -- that is, mode of formation of subject -- may give rise to different varieties and degrees of inter-relation among the component ideas. Therefore, in structuring and coextensive representation of subjects, either for designing a scheme for classification or for classifying, the different ideas forming components of subjects as well as the modes of formation of subject must be recognised. The following are some of the modes of formation of subject recognised so far (10). A brief description and illustrative examples of each of the modes are given in Appendix 2.

1 Lamination 1
2 Lamination 2
3 Loose assemblage 1
4 Loose assemblage 2
5 Loose assemblage 3
6 Fission
7 Fusion
8 Distillation
9 Partial comprehension
10 Subject bundle

Partial Comprehension and Subject Bundle are largely the result of developments in book production -- that is, in the universe of documents -- rather than being intrinsic to subjects. Research will bring to light new modes of formation of subject, just as Fusion, Distillation, and Subject Bundle were recognised only in the last five years. It is necessary to examine the modes of formation not only from the angle of the design of a scheme for classification but also from the angle of the subject specialists. A brief account of the views of Boulding, Tromp, Weiss, and others has been presented by Devadason. The concepts of interdisciplinary "hybrid sciences" and "border sciences" are noteworthy (11).

22 Dynamic Growth

At a given point of time, the classificationist may have to deal with only a finite universe of subjects. But if the scheme for classification he designs is to continue to be useful, he must recognise that the universe of subjects and the psychology of readers are ever-dynamic, ever-developing, and that the development of the universe of subjects is accelerating from time to time. An idea emerges when a knower inter-acts with a knowee or entity. An idea, therefore, is essentially the product of human intellect. Here we can think in terms of three correlated universes: (1) The universe of knowees consisting of known entities as well as unknown but knowable entities; (2) As and when the human intellect learns about an entity, the idea so generated is added to the universe of knowledge forming a subset of the universe of knowees or entities; and (3) Some of the ideas may get embodied in documents for purpose of communication and future use. This constitutes the universe of subjects, forming a subject of the universe of knowledge. Since the universe of knowees is practically limitless and the human intellect promises to be ever active, the flow of ideas from the universe of knowees to the universe of subjects through the universe of knowledge, would apparently be an ever-continuing process. Thus, the universe of subjects tends to become infinite.

221 Data on Dynamic Growth

Here are some data to illustrate the dynamic growth in the universe of subjects (12).

Table 1. Major New Ideas in Different Periods

Colon Class Number	Main Subject	Number of major new ideas and subjects during the period	
		1701-1900 (200 y)	1901-1950 (50 y)
B	Mathematics	24	22
C	Physics	21	70
D	Engineering	23	46
E	Chemistry	36	80
F	Technology	21	72
FZ	Biological sciences (excluding Genetics and Medicine)	17	20
G;6	Genetics	42	95
L	Medicine	21	81
X	Economics	36	32
Y	Sociology	63	49
	Total	304	567

The rate of output of new ideas during the first half of the present century is nearly six times that during the past two centuries.

Table 2. Acceleration of Development: In Different Subjects

SN	Subject	Year of		Time-lag between (c) and (d)
		Discovery of the principle	Development of device or use	
(a)	(b)	(c)	(d)	(e)
1	Internal Combustion engine	1710	1892	182
2	Photography ..	1727	1839	112
3	Aerial flight ..	1809	1903	94
4	Telephone ..	1823	1876	53
5	Gamma-ray ..	1896	1939	43

(a)	(b)		(c)	(d)	(e)
6	Atomic power release	..	1932	1945	13
7	Transistor	..	1940	1948	8
8	Laser	..	1958	1960	2

Table 3. Acceleration of Development: In a Subject-field

Marine Invention		First plan	Commercial success	Time-lag
Geared turbine	..	1929	1911	282
Steam boat	..	1661	1808	147
Propeller	..	1746	1836	90
Iron ship	..	1777	1829	52
Fixed contrapropeller	..	1864	1910	46
Diesel engine	..	1882	1903	21
Rotorship	..	1923	1926	3

Table 4. Accelerated growth and Increase in the Number of Papers

Year	Laser Technology	Approximate Number of Papers on Laser per year
1958	Laser structure proposed	1
1960	Ruby laser	20
1961	Helium-Neon laser	100
1962	Solid state laser	325
1963	GaAs laser	700
	Pulsed laser	
	Q Switching	
1964	Ion laser	1,000
	Coverage of 200 new wave-lengths	
1965	N_2- CO_2 high efficiency laser	1,200
1970		c. 1,500

Table 5. Inrease in the Number of Articles in Periodicals (13)

Colon Class Number	Subject	Number of articles	
		1960	1970
B	Mathematics	15,000	30,000
C	Physics	75,000	155,000
D	Engineering	155,000	275,000
	Civil	15,000	15,000
	Mechanical	10,000	20,000
	Electrical and Electronic	80,000	150,000
	Aero space	35,000	75,000
	Industrial	15,000	15,000
E	Chemistry	150,000	260,000
F,b	Metallurgy	35,000	50,000
G	Biology	150,000	290,000
GZ	Geosciences	91,000	158,000
J	Agriculture	150,000	260,000
L	Medicine	220,000	390,000
S	Psychology	15,000	30,000
	Other subjects	929,000	1,882,000
	Total	1,985,000	3,780,000

222 Kinds of Growth

The overall pattern of growth of ideas in a particular subject field may conform to a logistic curve. There may, however, be different kinds of internal "oscillations". Price (14) has mentioned four ways in which logistic growth may react to ceiling conditions. These are: (1) Escalation; (2) Loss of definition; (3) Divergent oscillation; and (4) Convergent oscillation. These oscillations present a picture of turbulent dynamism of the universe of subjects. A close examination of the properties of growth considered in the field of General Systems has also been found helpful (15).

223 Change of Strength of Bond

Inter-disciplinary research may lead to changes in the interrelation among the components in a subject going with a Basic Subject. It may also lead to change in the inter-relation between subjects going with different Basic Subjects. Usually, the change is towards increasing the strength of bond between the components of subjects (16) going with a Basic Subject, or the subjects going with different Basic Subjects as the case may be. Table 6 presents a few examples.

Table 6. Change of Strength of Bond

SN	Subject	Period in which it was deemed as			
		Complex Subject (Phase relation)	Compound Subject (Facet relation. Lamination)	Non-main Basic Subject (Fission)	Main Subject (Fusion)
1	Astrophysics		1933-68		1969
2	Chemical engineering	1933-51	1952-67	1968	1969
3	Biochemistry	1933-51	1952-57	1957-66	1967
4	Forestry	..	..	1933-68	1969
5	Public health	..	1933-68	..	1969
6	Medical jurisprudence	1933-49	1950-68	..	1969
7	Geopolitics	1933-68	..	..	1969

23 Consequence of the Growth of the Universe of Subjects

To the classificationist the attributes of growth of the universe of subjects mentioned in the preceding sections implies essentially the following:

1 The emergence of

11 New subjects,

12 New Main subjects of different varieties,

13 New Basic Subjects of different varieties going with one of the other of the Main Subjects,

14 New isolates for inclusion in any of the schedules of isolates, and

15 New schedules of isolates to be added to the existing schedules of isolates;

2 Sharpening of isolates in different ways; and

3 Increase in the average number of facets in subjects even in the book level.

231 Data on Basic Subjects

The impact of the accelerated growth of the universe of subjects during the past decade has been sensed while revising CC, Ed 6 (1960) to prepare its edition 7. Table 7 presents data about the new Basic Subjects etc in CC, Ed 7 (1972). It must be noted that data on Environmental Basic Subjects are not included in the table. This is because such Basic Subjects are to be derived using the Environment Device as and when required. A System Basic Subject is to be derived by using the Chronological Device as and when required. Therefore, the data included in the table covers only the enumerated Systems Basic Subjects which have been recognised at the book level (18).

Table 7. New Basic Subjects
(excluding Environmental Basic Subjects)

SN	Variety of Basic Subject, etc	CC, Ed 6	Newly added	CC, Ed 7
1	Main Subject (Total)	33	75	108
11	Main Subject (other than Fused and Distilled ones)	33	31	64
12	Fused Main Subject	-	28	28
13	Distilled Main Subject	-	16	16
2	Partial comprehension	9	22	31
3	Subject Bundle	-	9	9
4	Non-Main Basic Subject	347	228	575
		389	334	723

232 Increase in Intension

For convenience, the number of component ideas in a compound subject may be taken to measure its intension. The illustrative data presented in the following tables indicate that there is a tendency for the intension of even subjects embodied in whole books to increase with time.

Table 8. Incidence of Facets in Macro Subjects going with the Main Subject Medicine (825 books) (19)

Number of facets	1925		1951		1967	
	Number of Books	% of 25	Number of Books	% of 300	Number of Books	% of 500
1	4	16	10	3.3	10	2.0
2	9	36	71	23.7	65	13.0
3	4	16	93	31.0	170	34.0
4	7	28	95	31.7	171	34.2
5	1	4	31	10.3	75	15.0
6	-	-	-	-	9	1.8

Table 9. Incidence of Facets in Macro Subjects going with the Host Subject Economics of Education (200 books)

Number of facets	1931-60	1961-68
	% of 50 books	% of 150 books
2	15.0	10.0
3	35.0	24.0
4	31.0	40.0
5	11.0	20.0
6	3.0	5.0
7	0.5	1.0

233 Intension in Macro- and Micro-Subjects

A macro subject is usually embodied in a whole book. A micro subject is usually embodied in a section of a book, an article in a periodical, a technical report, a patent, etc. However, a subject embodied only as an article in periodicals today may get embodied in whole book also in due course. For example, "Laser" was the subject of an article in 1958. The first monograph on it was published within two years thereafter. In the last five years, the average number of books on Laser published per year is ten. Such features of the universe of subjects imply that differentiation between macro subject and micro subject on the basis of the number of component ideas incident may not be precise or even useful for the purpose of designing and developing a scheme for classification. Table 10 presents data derived on the basis of facet analysis of the subjects of 200 books and 100 articles on Economics of Education.

Table 10. Number of Facets in Books and Articles

Number of facets	Incidence (as %) in	
	Book	Article
2	13.0	3.0
3	30.0	25.0
4	36.0	34.0
5	16.0	24.0
6	4.5	11.5
7	0.5	2.5

The mode is four facets in both the cases. Nearly 83 per cent of the documents have 3 to 5 facets in the book level as well as article level. There is a higher percentage of books with two facets as compared with the articles; and there is a higher percentage of articles with six or more facets as compared with the books.

24 Tendency to Become Continuum

The incessant intellectual probe not only continually adds new ideas and new subjects to the universe of subjects but existing gaps in our knowledge also get filled up. The gap may be of any "size". This implies that each isolate idea would claim a particular place, preferred by the majority of specialist readers, in the subject-field concerned; and each subject would claim a particular place in the sequence of subjects going with a particular Basic Subject and also in the totality of the universe of compound subjects. The position of the subject is to be determined by the mutual filiation of the subjects in the universe of compound subjects and according to the preference of the majority of the specialist readers.

25 Implication to Classification

In the design and development of a scheme for classification, all this implies the need for:

1 Extrapolation in any Chain -- be it a Chain of Basic Subjects or of isolates;

2 Interpolating in any Chain -- be it in a Chain of Basic Subjects or of isolates; and

3 Interpolating as well as extrapolating of new schedules of isolates among the already existing ones.

Thus, the design of a scheme for classification, though intended at any one time for the then existing finite universe of subjects, it must be capable of accomodating each of the new ideas and subjects as and when they emerge and get embodied in documents, without the basic structure of the scheme being damaged to any appreciable extent. In fact the classificationist has to anticipate the likely developments, at least in the near future, in the universe of subjects.

3 FACTORS AFFECTING THE MENTAL SET OF READERS

31 Enumeration

Experience gained in doing reference service to readers -- that is, helping each reader to find all the documents meeting his exact needs at the moment -- and also the findings of special studies and surveys of "user habits", lead to the recognition of the following factors as affecting the mental set of a reader while searching for his documents:

1 The universe of subjects is ever-growing. It throws forth new subjects continuously. It is multidimensional;

2 Most of the subjects are compound ones having a sequence of constituent kernel ideas -- that is, facets;

3 The particular fundamental pattern of linear sequence among the facets of any compound subject that is helpful to the majority of readers (See Sec 82);

4 Document finding is facilitated, if the main entries of the documents on a stated small subject-range are presented to the reader in a sequence helpful to the majority of readers, and have Feature Headings of each subject in the range arranged according to Facet Syntax (20); and

5 While aiding the reader in formulating his subject-interest at the moment it is helpful to facet analyse his subject by the same method as the one used in the facet analysis of the subjects of documents (See also sections 15 and its subdivisions).

32 Multiplicity of Facet in Subject of Reader's Query

One of the results of extreme specialisation in narrow subject fields is the atomisation of the research on and development and production of, commodities. Design and production engineers seek information about a product giving a large number of specifications. For example, Hutton (21) has pointed out that is not unusual for architects and design engineers to require information on products with specifications such as follows:

Metal Window with the following specifications:

> Double hung vertical sliding sash; 5 feet 8 inch x 4 feet wide; made of aluminium alloy HE9; satin anodised finish to grade AA35 with triple lacquer protection; gear operated; plastic coated operating handle; suitable for double glazing units.

Today subject specification also often contains a large number of facets. For example, the facet analysis of the subjects of 323 queries put by specialists in the fields of high speed aerodynamics and aeronautical engineering, showed that the average number of facets incident in a subject of a query was 5.4, with a maximum of 12 and minimum of 2. Also, the average number of kernel ideas per subject was 8, with a maximum of 18 and minimum of 3.

4 OBSOLESCENCE OF SCHEME FOR CLASSIFICATION

The result of the impact of the attributes of the universe of subjects, the specificity of subjects of reader's query, and of the psychology of readers using document finding systems, on schemes for classification, is a high rate of obsolescence of the latter. The rate and mode of development of the universe of subjects further shows that the obsolescence rate of schemes for classification will increase in future.

5 NEED FOR A DYNAMIC GENERAL THEORY OF LIBRARY CLASSIFICATION

51 Efficiency of Document Finding System

If a scheme for classification forms an essential element built into a document finding system, the efficiency of the latter will depend upon the efficiency of the scheme for classification in keeping step with the incessant growth of the universe of subjects and also with the variation, from time to time, of the subject approach of readers. To secure this efficiency, the scheme for classification should be based on and keep step with, the developments in the theory of library classification.

52 Arrangement of Compound Subjects

The number of compound subjects going with a particular Basic Subject is large. The number of compound subjects going with different Basic Subjects is far too large to be arranged helpfully and consistently without the aid of explicitly stated guiding principles. For, such a universe is manifold multi-dimensional. Classifying is equivalent to transforming its _n_-dimensional configuration into one-dimensional or linear configuration. The difficult problem of "invariant" arises in such a transformation or mapping (22).

521 Helpful Sequence

An arrangement of the compound subjects going with a particular Basic Subject in a sequence helpful to the majority of the specialist readers may be said to result in an APUPA pattern -- that is, Alien-, Penumbral-, Umbral-, Penumbral-, Alien- subjects. For example, the sequence of the following subjects is deemed to be in an APUPA pattern: Economics, Banking, Commercial Bank, Credit, Loan against mortgage, Investment, Bond, Yield on Bond. It is impossible to secure APUPA pattern everywhere in the entire range of the compound subjects. For, it requires keeping invariant every immediate-neighbourhood-relation among all the subjects while mapping the n-dimensional universe of compound subjects in a line.

522 Consistency of Pattern in the Sequence of Compound Subjects

Thus, in mapping the multidimensional universe of compound subjects going with a Basic Subject, the classificationist is obliged to keep invariant only one of the many immediate-neighbourhood-relations found in an *n*-dimensional space. The question is as to whose immediate-neighbourhood-relation should be kept invariant, or which point should be of remove 1, which should be of remove 2, and so on. These decisions cannot be left to the conjectures of classificationists. For, it will lead to inconsistency in the pattern of sequence. But, consistency in the pattern of sequence of the compound subjects going with all the different Basic Subjects would indeed be greatly helpful to the reader, the classifier, and the classificationist.

53 Genesis of Theory of Library Classification

The need for a general theory of library classification to guide the work of designing a scheme for classification and for classifying arises out of: (1) The difficulty of choosing a particular succession of immediate-neighbourhood-relation of remove 1, remove 2, and so on; and (2) Securing a similar pattern of arrangement in the very large number of compound subjects going with different Basic Subjects.

6 GENERAL THEORY OF LIBRARY CLASSIFICATION

61 Early Descriptive Theory

For a few decades after Dewey's pioneering work, library service as a whole was largely a matter of trial and error and rule of thumb. Flair and experience helped the librarian whenever any difficulty arose. In due course, it was realised that library service should keep step with the changes in the universe of subjects and in the reader's requirements. It was also beginning to be realised that classification by subject should form the basis for efficient library service and therefore, classification should also keep step with the changes in the universe of subjects and

COS
202

in the reader's requirements. In the early part of this century on the basis of a comparative study of the existing schemes for classification, E C Richardson (23), W C Berwick Sayers (24), and Henry E Bliss (25) attempted to formulate some empirical principles and a theory of library classification. But this theory was largely descriptive. It was neither general nor did it have internal dynamism. The difference in the foundation of UDC and DC was not appreciable. There were no other schemes for library classification with distinctly different kinds of foundation for comparative study. Therefore, the probability for the emergence of a general dynamic theory of library classification was low.

62 Current Dynamic Theory

The publication of Colon Classification (1933), which was based on a totally different kind of foundation, provided material for a truly comparative study of classification, and therefore, to the development of a dynamic general theory of classification. Another favourable factor was the formulation of the Five Laws of Library Science by Ranganathan in 1928 and their publication in 1931 (26). These laws provided a unified view of library science and library service as a whole, and formed the basis for the derivation of guiding principles by deduction for everything connected with library service. They could be used as the basis for a comparative study and evaluation of document finding systems. With respect to classification, the laws were of help in driving the foundation piles of the theory of classification to a greater depth, so as to base them on a firm rock as it were. Sets of postulates, canons, and principles for library classification could be formulated and explicitly stated. The first formulation of the resulting theory was embodied in the Prolegomena to library classification (1937). On the basis of the findings of research on the attributes of the universe of subjects and of the demands of readers, this theory has been continuously refined, thereby steadily increasing its internal dynamism. Edition 3 of the Prolegomena (1967) consolidated the results obtained till about 1966. Thanks to the team work in DRTC, further improvements are being frequently reported in the quarterly Library science with a slant to documentation (1964-) and in the

Annual Seminar (DRTC), Papers and proceedings (1963-). This is perhaps the only dynamic general theory of classification now current. Its latest version has highlighted the great value of a freely faceted analytico synthetic classifcation over other models of classification as an aid in documentation. We shall use this theory as the reference point in studying the compatibility between UDC and CC later in this paper.

621 Normative Principles

The current general theory of library classification is based on a hierarchy of normative principles. The Prolegomena (1967) discusses 91 of them referred to in the following table.

Table 11. Normative Principles

Name of the Set of Normative Principles	Number of Principles	Reference to Prolegomena, Ed 3
General Laws	6	Chap DC to DH
Laws of Library Science	5	Sec DB1 to DB5
Canons of Classification for work in the Idea Plane	15	Part E
Canons of Classification for work in the Verbal Plane	4	Part G
Canons of Classification for work in the Notational Plane	21	Parts J to L
Canons for Book Classification	3	Sec UB2, VB2 and VC3
Postulates of Classification	17	Part R
Principles for Facet Sequence	4	Sec RM1, RN1 to RN3
Principles for Helpful Sequence	16	Part P
	91	

7 CC AND UDC FROM THE ANGLE OF THE GENERAL THEORY

71 Colon Classification

Edition 1 of CC (1933) was not based on any objective explicitly stated theory of classification. Probably, some unexpressed theory was guiding its design from the subconscious level even as it should have happened to DC. From Ed 2 (1939) onwards, CC has been guided by and attempting to objectively conform to, the general theory of library classification embodied in the Prolegomena and its refinements.

72 Universal Decimal Classification

During the past two decades a few guiding principles have been formulated for guiding the development of UDC (27). But these do not constitute a comprehensive and deeply founded theory. Nor are they supported by explicitly stated sets of more general normative principles. Hence the theory by which UDC is being guided lacks internal dynamism. The theory mainly deals with the multitudinous phenomenal level than with the deeper near seminal level of which the phenomenal levels are diverse manifestations. Even so a comparative study shows that at the subjective level some unexpected deeper general theory might be guiding UDC also. For, some of the elements of this theory agrees with the objective theory of the Prolegomena. The following tables show the extent of conformity of CC and UDC to the current dynamic general theory of classification in respect of the idea plane and the notational plane.

721 Conformity of CC and UDC in the Idea Plane

Points of Comparison	CC	UDC
Separation of planes of work into Idea, Verbal, and Notational Planes, making the first paramount and making the last implement the findings of the first.	Takes maximum advantage of this guidance of the theory.	There is occasional evidence of some distinction being made at subconscious level. But the paramountcy of the Idea Plane and the obligation of the Notational Plane are not always carried out. The Notational Plane often inhibits the work in the Idea Plane.

Points of Comparison	CC	UDC
Fundamental pattern in the structure of the compound subjects going with different basic subjects.	Uniformally follows the postulates of the five Fundamental Categories and provides for Rounds and Levels of facets as provided in the general theory.	Not all the facets and phase relations are distinctly separated due to the inhibition of the work in the Idea Plane by the shortcomings of th Notational Plane and due also to its DC core having many compound subjects.
Fundamental pattern in the sequence of the facets of the compound subjects going with different Basic Subjects.	Uniformally follows the postulates and the principles provided in the general theory.	Secures only partially.
Uniqueness of class number for a subject.	This is secured by strict adherence to the general theory.	There is deliberate provision for alternative class numbers for a Subject.

722 Conformity of CC and UDC in the Notational Plane

Points of Comparison	CC	UDC
Length of base	60 for use in the array of main subjects. made up of 23 Roman smalls (excluding i, 1, and o), 10 Indo-Arabic numerals, 26 Roman capitals and the Greek letter Δ. In the other arrays, 58 only, because I and O are omitted.	10 made up of the Indo-Arabic numerals only

Points of comparison	CC	UDC
Extrapolation	Secured by Sector Notation	Secured by Octave Notation after the acceptance of the recommendation made by FID/CA in 1955.
Interpolation in Array	Secured, to any extent wanted, by the use of Empty Digits (Sector Notation), Emptying Digits. Sector Notation alone gives the capacity of 1166 in an array in which no array isolate number has more than three digits. The capacity can be increased indefinitely by the use of Emptying and of Empty-Emptying Digits.	Use of brackets.
Extrapolation in Chain	Secured by using decimal fraction notation.	Secured by using decimal fraction notation.
Hospitality among facets	Secured to any extent wanted by following the postulates of Rounds and Levels provided by the theory and by the use of a distinctive Indicator Digit for each Fundamental Category	By repeated use of Colon or square brackets and the use of other indicator digits -- such as, circular brackets, double inverted comma, hyphen, and 0.

Points of Comparison	CC	UDC
Mnemonics	Satisfied the Canon of Scheduled, Systematic, and Seminal Mnemonics.	Satisfies the Canon of Scheduled Mnemonics to a limited extent.
Expressiveness of Class Number	Secured by the use of different Indicator Digits to indicate the nature of the succeeding facet.	Secured in respect of Space and Time facets only by the use of distinctive Indicator Digits, but not in the other cases.

73 CC and UDC in Practice

Viewed from the angle of the current general theory of classification there appears to be much incompatibility between CC-in-theory and UDC-in-theory. But a comparative study of the two schemes in practice -- that is, in classifying -- has shown that in most cases UDC numbers can be constructed according to the general theory of classification. Thus UDC-in-Practice admits of being so used as to reduce its incompatibility with CC-in-Practice to almost a vanishing point. This is shown in the following section by classifying a few illustrative examples of sbbjects by the two schemes. The subjects are classified according to the Method of Postulates (28). Some of the steps of the method, are not given as they are not necessary in the comparative study. In constructing the class number according to UDC the prescriptions in the Guide to the Universal Decimal Classification (1963) has also been used wherever helpful.

731 Examples of Classification

Abbreviations used:

(BS) = Basic Subject

[1M1] = Matter Facet, Round 1, Level 1

[1M2] = Matter Facet, Round 1, Level 2
[1P1] = Personality Facet, Round 1, Level 1
[S1] = Space Facet, Level 1
[T1] = Time Facet, Level 1

Example 1.

Step 0	Raw Title	Audio-visual technique in teaching. Note.- 1 A perusal of the document indicates that it gives a descriptive account of the state of high school teaching of carpentry using audiovisual aids in Japan in the 1950s. 2 The subject is primarily of interest to one belonging to the profession of teaching. It has less interest, if at all, to one belonging to the craft of carpentry. Therefore, Law 2 of Library Science would prefer it to be taken along with the Basic Subject "Education".
Step 1	Expressive Title	Audio-visual teaching technique for carpentry in high school level education in Japan in the 1950s.
Step 2	Kernel Title	Audio-visual. Teaching technique in Carpentry. High School. Education. Japan. 1950s.
Step 3	Analysed Title	Audio-visual [1M2]. Teaching technique in Carpentry [1M2]. High school [1P1]. Education (BS). Japan [S1]. 1950s [T1].
Step 4	Transformed Title	Education (BS). High school level [1P1]. Teaching technique in carpentry [1M1]. Audio-visual [1M2]. Japan [S1]. 1950s [T1].

Step 5	Title in Standard Terms	According to UDC: Education (BS). Secondary schools [1P1]. Method of instruction (in) carpentry [1M1]. Oral and visual method [1M2]. Japan [S1]. 1950s [T1]. According to CC: Education (BS). Secondary [1P1]. Teaching technique (in) carpentry [1M1]. Audio-visual [1M2]. Japan [S1]. 1950s [T1].
Step 6	Title in Kernel Numbers	According to UDC: 37 (BS). 35 [1P1]. 371.33 694.2 [1M1]. 33 [1M2]. 520 [S1]. 195 [T1]. According to CC: T (BS). 2 [1P1]. 3 (M2) [1M1]. 01 [1M2]. 42 [S1]. N6 [T1].
Step 7	Class Number	According to UDC: 373.5:371.33:694.2(520)"195" According to CC: T,2;3(M2);01.42'N5

Annotation

In this case, UDC is compatible with CC both in theory and in practice. For, they conform to the same general theory of classification, recognise the same kernel ideas and the same sequence for them, in the subject.

Example 2

Step 0	Raw Title	Hereditary diseases of the eye
Step 1	Expressive Title	In Medicine, hereditary diseases of the eye.
Step 2	Kernel Title	Medicine. Hereditary disease. Eye.
..		.. .. ..
Step 4	Transformed Title	Medicine (BS), Eye [1P1]. Hereditary disease [1M1]

Step 6	Title in Kernel Number	**According to** UDC: 61 (BS): 611.84 [1P1]: 616-056.7 [1M1]. **According to** CC : L(BS),185[1P1];4G[1M1]
Step 7	Class Number	**According to** UDC: 61:611.84:616-056.7 **or** 61 [611.84] [616-056.7] **According to** CC: L,185;4G

Annotation

As in Example 1, in this case also, CC number and UDC number are equivalent and coextensive with the subject of the document. Therefore, they are compatible with each other. However, to bring UDC in conformity with theory, a compound subject has to be used as if it were an isolate idea. Further, in theory, UDC gives freedom to arrange the coloned subject numbers in any sequence. This violates the postulates of a unique structural pattern for a subject. To avoid this violation, square brackets can be used. But it can lead to the use of one and the same device for two different kinds of divisions -- namely, array division and facet division. The sequence chosen in the above example conforms to the general theory of classification. Thus, although CC and UDC are not necessarily compatible in theory, compatibility has been secured in practice.

Example 3

Step 0	Raw Title	Genetics for opthalmologists
Step 1	Expressive Title	In Biology, genetics is expounded with a bias to medicine of the eye.
Step 2	Kernel Title	Biology. Genetics bias to Medicine. Eye.
	..	.. .. ..
Step 4	Transformed Title	Biology (BS). Genetics [1M1]. (Phase relation) bias Medicine (BS). Eye [1P1].

Step 6	Title in Kernel Numbers	According to UDC: 57(BS) 5 [1M1]: 617 (BS) 7 [1P1].
		According to CC: G(BS); 6 [1M1]&bL (BS), 185 [1P1]
Step 7	Class Number	According to UDC: 575:617.7
		According to CC: G;6&bL,185

Annotation

In this case, UDC in practice is not compatible with CC in practice. Nor does UDC conform to the reference point -- that is, the general theory of classification. For, the UDC number for the subject of the biasing phase represents the narrow one of "surgical diseases of the eye" instead of the more extensive subject, namely "Eye in all its properties" as mentioned in the subject of the document classified.

However, this incompatibility can be corrected according to the principles given in the Guide to the Universal Decimal Classification (1963). Accordingly, the UDC number will be

57;575:61;611.84.

This shows that UDC in practice, by adopting a suitable policy, can be made compatible with the general theory of library classification.

8 SYSTEMS APPROACH

81 Adaptability

In another paper (15) it has been shown that the Generalised Facet Structure based on the General Theory of Classification of Ranganathan is a model for structuring in a helpful way any subject in the universe of subjects. It has been further pointed out in the same paper that such a structuring endows the representation of subjects with attributes of a "System" as generally conceived and used in General Systems Theory developed by Von Bertelanffy, Hall and Fagen, Boulding, Khailov and others.

A classification system which patterns a subject in this way should also enable the search and selection of documents or their surrogates

1 To any predetermined degree of relevence;
2 To any predetermined degree of recall;
3 With any predetermined optimum expeditiousness.

The classification system must continue to provide such facility in the document finding system in which it is used even as it develops, adopts, and evolves to meet the dynamic changes in the universe of subjects and requirements of readers, without its basic fundamental characteristics being changed. To endow the classification system with such a resilience and self-perpetuating quality, it should have a built-in mechanism. Homoeostacis helps a system maintain equilibrium in different contexts, if the opposing/disturbing force is not strong enough to shake the foundation or basic structure of the system -- that is, endows it with a self-organising quality. For a classification system such as UDC the built-in mechanism can be in the form of guiding principles (See Sec 621) which can form the basis not merely for classifying but more importantly for the design and development of the classification system.

82 Illustrative Example

In order to illustrate that the available set of postulates and principles of the General Theory of Library Classification can indeed provide such a built-in mechanism, the Facet Syntax in a Compound Subject is briefly discussed below.

821 Absolute Syntax of Ideas

The sequence in which the component ideas of compound subjects going with a Basic Subject usually arrange themselves in the mind of the majority of normal intellectuals may be called Absolute Syntax of Ideas among intellectuals. It is conjectured that such a syntax of ideas exists. It may not coincide with the linguistic syntax -- that is, the syntax of words in all languages. Investigation by a team consisting of reference librarians, classificationist specialists in linguistics, epistemology, psychology, anthropology, and statistics has been suggested to examine the validity of the conjecture (20). This is yet to happen. In the meanwhile, however, several helpful results in the theory and practice of classification have been derived on the basis of this postulate about Absolute Syntax. The possible basis for the conjecture is mentioned in the succeeding section.

822 Biological Basis

An Idea is generally a product of intellectual activity. Intellectual activity is known to be controlled by the brain. There is considerable similarity in the structure and, therefore, the functioning of the brain in most of the normal human beings. In other words, a majority of the normal human beings have more or less a similar mode of thinking and learning -- that is, in forming ideas and in combining them to build subjects.

It is further believed that biologically man has not changed to any appreciable extent since he became *Homo sapiens* many thousands of years ago. The educationist William H Kilpatrick writes: "It is generally believed that man has not developed biologically in any significant manner or degree since he became *Homo sapiens*; culturally, yes — greatly so; biologically as regards mind, no, not so far as we can tell" (29). Or again, as Bertrand Russel writes: "There had been a time when there were biological improvements in brain capacity with corresponding advande in genetic capacity. But that time ceased some 500,000 years ago" (30).

823 Inference

From the above discussion we may infer that:

1 The mode of thinking/learning among normal intellectuals is similar;

2 The mode of thinking/learning among normal intellectuals has remained basically in the same pattern for several thousands of years; and for this reason,

3 The probability of a sudden change -- that is, a mutation -- in the mode of thinking/learning of a majority of normal intellectuals in the near future is quite low.

824 **Facet Sequence**

It has been suggested that the sequence of the component ideas in compound subjects — that is, the facet syntax — should parallel the Absolute Syntax of Ideas. For, if we can secure a facet syntax parallel to the Absolute Syntax, then the pattern of sequence of component ideas in compound subjects will be

1 Helpful to a majority of normal intellectuals; and

2 Consistent and more or less stable and hence continue to be helpful to a majority of intellectuals at least in the near future.

The sequence of component ideas thus derived is reader-requirement-based. Incidentally, the aberration likely to arise by using the verbal plane in structuring and naming subjects, due to the variation in linguistic syntax, is reduced to a minimum.

825 **In Practice**

A number of principles for securing a sequence of ideas deemed to be helpful to a majority of readers have been formulated and explicitly stated. These have been used for several years now in designing schemes for classification of subjects, in classifying, and in other contexts (31). One of these Principles of Helpful Sequence is the Wall-Picture Principle. It states: "If two facets A and B of a subject are such that the concept behind B will not be operative unless the concept behind A is conceded, even as a mural picture is not possible unless the Wall exists to draw upon, the facet A should precede the facet B" (32). It has been found that the Wall-Picture Principle

1 Is the most ubiquitous in its application; and

2 Secures a sequence of the component ideas in subject that is helpful to a majority of specialists.

It may, therefore, be said that the facet syntax resulting from the application of the Wall-Picture Principle parallels the Absolute Syntax of Ideas. It has been further found that the other Principles of Helpful Sequence are corollaries of the Wall-Picture Principle. Therefore, the sequence of ideas resulting from the application of any of the other Principles for Helpful Sequence will conform to that derivable by application of the Wall-Picture Principle itself. In a particular context, the use of one of the other principles may be more convenient than to work with the Wall-Picture Principle.

83 Bridge Language

Such a controlled classificatory language of ordinal numbers is not only independent of the linguistic syntax, but can conveniently form a "bridge language" or "switching point" from one indexing language to another in representing a subject. For example, the Class Number can be translated into a representation of the subject in the verbal plane in different natural languages.

9 COMPATIBILITY BETWEEN CLASSIFICATION AND COMPUTER-AIDED DOCUMENT FINDING SYSTEM

91 Classification and Computer-Aided System

The vital role of classification in document finding systems aided by electronic machinery has been reviewed by Bhattacharyya (33). The psychology of the reader with respect to his document requirements is the same whether he uses a conventional catalogue or a computer-aided document finding system. Therefore, an important objective of experiments in the use of a computer-aided document finding system has been to find out how far it can simulate the conventional document-finding system in satisfying the psychology of the reader. It is necessary for the librarian to ask the electronic engineer to design the machine to satisfy his specifications in respect of its suitability to meet the psychology of the reader. Experience with document finding systems using general purpose computers during the last decade, particularly in USA, USSR, and more recently in UK, supports this demand of the library profession. Further, it is being realised in an increasing measure that:-

1 The use of electronic system does not obviate the use of classification in it;

2 The potentiality of use of an electronic system could be better exploited, if

21 The subject of a document is put into the memory of the system in a correctly faceted form coextensive with the subject,

22 The subject of the question put by reader is also similarly facet-analysed;

23 The components found by facet analysis are arranged in a fundamental pattern of helpful sequence, and

24 The entries for the documents selected by the machine as relevant to the reader's question are displayed in a classified sequence with necessary Feature Headings to help browsing at each stage of the trilogue, so that a more precise statement of the subject sought and selection of the correct entries by the reader are facilitated (See Sec 153 above)

92 Compatibility of CC

A few years ago, it was thought that CC, with its notational system having over 71 digits, was too complicated for computers. But the computers of today can handle even a much larger number of digits -- as many as 120 and even more. In fact, experiments at the DRTC (Bangalore) during the last two years have shown that not only is CC compatible with computer-aided document finding system, but the scheme offers certain special advantages.

93 Compatibility of UDC

There are already several computer-aided document finding systems using UDC. Experiments are being done at several centres to improve the efficiency of the use of the scheme in such systems. Thus, UDC is also compatible with computer-aided system. See for example FID/CR report 9 (1970).

94 Advantage of Using a Theory-based Classification System in Computer-Aided Systems

In the discussion in Sec 721 and 722, the possibility for internal inconsistencies in UDC was pointed out. The main reason for this is that the UDC and its developments are not objectively based on a explicitly stated dynamic theory of library classification. On the other hand, a theory-based classification system could avoid such internal inconsistencies. Therefore, it will havethe capacity and resilience for

1 Receiving any new subject in the correct helpful place among the already existing subjects, as determined by the Idea Plane;

2 Keeping pace with the refinements in the General Theory of Classification;

3 Self-perpetuation;

4 Providing a unique coextensive class number to a subject, however deep, so as to place it in the correct helpful place among the already existing subjects; and

5 Meeting new demands of readers.

D J Foskett has drawn attention to the particular advantage offered by the CC methodology in computer-aided document finding systems (35). Also, Landgridge has pointed out that some of the problems met with in computer-aided document finding system using UDC get resolved or are not present when CC is used (36). Thus, in the balance, whether it be in a conventional document finding system or in a computer-aided one, CC has advantages over UDC. The provision of a coextensive unique representation of a subject is particularly important to minimise noise in the search and selection of documents. The following examples are illustrative.

Consider the subjects resulting from a combination of the ideas represented by the following terms:

Production engineering
Motor car
Diesel engine

At least two different subjects can be built from a combination of these ideas:

1 Production engineering of motor car with diesel engine; and

2 Production engineering of diesel engine for motor car.

Subject 1 is of predominant interest and normally falls within the field of specialisation of Motor car production engineers. Subject 2, on the other hand, is of predominant interest and normally falls within the field of specialisation of Diesel engine production engineers. The structuring of the subject in the two cases will be different although the combining constituent ideas are similar.

Similarly, consider the subjects formed by a combination of the ideas represented by the following terms:

Missile
Destruction
Underground
Dump

At least two different subjects can rise:

1 Destruction of missile by underground dump; and
2 Destruction of underground missile dump.

95 Universal Classification System

951 FID and UDC

The FID has been fostering the development and the use of UDC. The advantage of having a universal classification system used throughout the world is obviously well appreciated. Efforts are being made to develop UDC into such a universal system for classification. In another paper (37) we have examined the implications of this in some detail. We have sorted out and emphasised the intrinsic and essential attributes of an efficient universal classification system. We have also examined whether UDC can be developed into a universal classification system.

96 Developing UDC

On theoretical considerations and from practical experience, it will be conceded that a dynamic general theory of library classification should form the foundation and provide the guidelines for developing and fitting a scheme for classification adequate to meet the challenges of an ever-dynamic universe of subjects and the reader's psychology. If this is conceded, then there is no reason why UDC should not be based and guided by such a dynamic theory. Over the past three decades such a theory of library classification has been developed and continuously refined on the basis of the findings of the studies of the attributes of the universe of subjects and psychology of readers. Therefore, there is no need to start *de novo* developing a theory for UDC. Conservation of the research potential in library science requires that we make use of the available theory and through further research improve upon it whenever necessary. There is, perhaps, still a misunderstanding in some quarters that the current dynamic general theory of classification mentioned in this paper and embodied in the Prolegomena is for CC only. This impression may be due to the fact that the author of this theory and the author of CC happen to be one and the same person. From personal experience I know that Ranganathan evaluates CC impartially on the basis of the theory as if he were not the author of the former. However, it is true that CC has been continuously attempting to implement and approximate to the prescriptions of the theory, the theory itself is a general one being based on the attributes of the universe of subjects and psychology of readers. It is hoped that those concerned with the development of UDC will consider this seriously and fulfil what Donker Duyvis wished for (38).

97 Recommendation

On the basis of the discussion in this paper the following recommendations are made:

1 FID/CCC and FID/CR should examine the existing General Theory of Classification with a view to basing the design and development of UDC on it.

2 FID/CCC should produce and keep updated periodically a "Basic edition" of UDC sufficient to meet the demands of classification at the book level.

3 FID/CCC may sponsor and coordinate the development of depth classification version of UDC schemes for a particular subject or group of subjects on demand and in collaboration with an institution. What develops in this way can be indicated in the Basic Edition of UDC and in other communications of FID.

4 The problems in theory and application of theory arising in designing and developing of depth versions of UDC may be considered by FID/CR groups.

5 FID/CC should have a sufficient number of competent personnel for the above purposes.

The adoption by FID/CCC of a policy such as that mentioned above is more likely to help the development of UDC into a universal classification scheme and a switching language.

APPENDIX 1: TERMINOLOGY

(See Section 013 of the text).

1 Idea.- The product of thinking, reflecting, imagining, etc obtained by intellect by integrating with the aid of logic, a selection from the apperception mass, and/or what is directly apprehended by intuition, and deposited in the memory.

2 Subject.- An organised or systematised body of ideas whose extension and intension are likely to fall coherently within the field of interest and comfortably within the intellectual competence and the field of inevitable specialisation of a normal person.

3 Isolate Idea.- An idea or idea-complex fit to form a component of a subject, but not by itself fit to be deemed to be a subject. **Example:** "Bacteria" denotes an isolate idea. It is not, by itself, fit to be a subject. But it is fit to be a component of subjects, such as Biology of bacteria, and Diseases of human body caused by bacteria.

31 Compound Isolate.- An isolate idea in an isolate facet of a compound subject formed by combining two or more isolate ideas taken from one and the same schedule of isolates. (See Appendix 3).

32 Special Component of Compound Isolate.- An idea which is not by itself a subject or an isolate, but which can be used as a subcomponent to be attached to an isolate in order to form a compound isolate. **Example:** In the schedule for Linguistics, "Variant", "Slang", and "Dialect" of a language are enumerated as Special Components.

In the schedule for Space Isolates, "Physical features of land," "Orientation divisions", and "Empire divisions" are enumerated as Special Components.

4 Array and Chain.- The addition of a characteristic to a basic subject idea or an isolate idea leads to the addition of an array. The addition of an array implies the addition of a further subdivision of a basic subject or an isolate idea, as the case may be. Any isolate idea taken along with its succession of components is also an isolate idea. The succession of the isolate ideas from the first to the last reached in this process is denoted by the term "Chain of isolate ideas". In a similar manner, a chain may be

formed in respect of basic subject idea. Example of a chain of isolate ideas: Asia, India, Mysore State, Bangalore District, and Bangalore form a chain of isolate ideas. Example of a chain of basic subject ideas: Mathematics, Algebra, and Determinants form a chain of basic subject ideas.

5 Simple Subject.- Basic subject taken by itself. Example: Geometry, Medicine, Metaphysics.

6 Compound Subject.- A subject with a basic subject and one or more isolate ideas as components. Example: In the subject Absorption spectrum of electro-magnetic radiation in radiation physics, "Electromagnetic radiation" and "Absorption spectrum" are each an isolate idea; "Radiation physics" is a Basic Subject.

7 Facet.- Generic term to denote (1) the basic subject alone in a simple subject, and (2) either the basic subject component or the isolate component in a compound subject.

71 Compound Facet.- A facet having a compound basic subject or a compound isolate.

8 Main Subject.- A subject included in the schedule of Main Subjects. The presumption is that no Main Subject can be expressed in terms of the other Main Subjects in the schedule.

91 Non-Main Basic Subject.- A Canonical, or Specials, or System subdivision of a Main Subject.

APPENDIX 2 : MODES OF FORMATION OF SUBJECT

(See Section 21 of the text).

In the Universe of Subjects

1 Lamination 1.- In this mode, one or more isolates (I) are laminated over a Basic Subject (BS). It gives rise to a Compound Subject.

Example:

1 Radiation physics (BS). Light (I). Spectrum (I).
2 Medicine (BS). Eye (I). Disease (I). Treatment (I). Drug (I
3 Banking (BS). Commercial bank (I). Deposit creation (I). India (I). 1950-1965 (I).

2 Lamination 2.- In this mode, either

1 Two or more species of non-Main Basic Subjects going with the same Main Basic Subject are laminated over one another, giving rise to a Compound Basic Subject; or

2 Two or more isolates from the same schedule of isolates are laminated over one another, giving rise to a Compound Isolate.

Example:

For Compound Basic Subject:

1 Lamination of the Systems Basic Subject "Quantum Theory in Physics" with the Canonical Basic Subject "Radiation Physics", giving rise to the Compound Basic Subject "Quantum Theory of Radiatio

2 Lamination of the Systems Basic Subject "Ayurveda" in Medicine, the Environmental Basic Subject "Tropical Medicine", and the Specials Basic Subject "Child Medicine", giving rise to the Compound Basic Subject "Ayurvedic Tropical Child Medicine".

For Compound Isolate

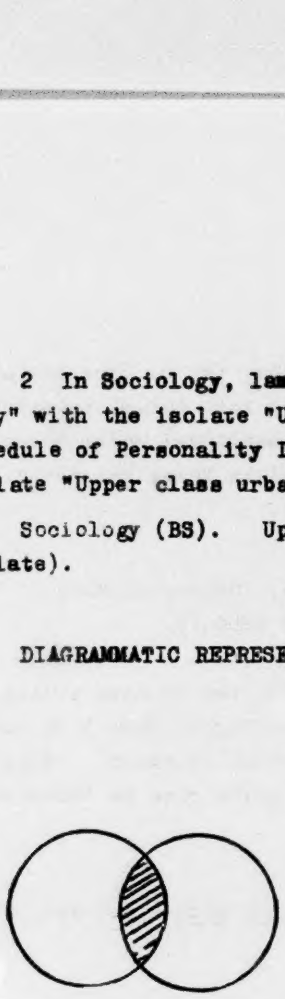
1 In Medicine, lamination of the isolate "Vein" with the isolate "Arm", both taken from the schedule of Personality Isolates, giving rise to the Compound Isolate "Veins of the Arm". Thus

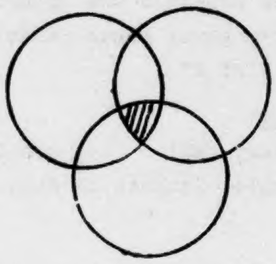
Medicine (BS). Vein-Arm (Compound Isolate).

2 In Sociology, lamination of the isolate "Upper class community" with the isolate "Urban community", both taken from the schedule of Personality Isolates in CC, giving rise to the Compound Isolate "Upper class urban community". Thus,

Sociology (BS). Upper class-Urban community (Compound isolate).

DIAGRAMMATIC REPRESENTATION OF LAMINATION IN GENERAL





3 Loose Assemblage 1.- In this mode, two or more subjects -- simple or compound -- are studied in their mutual relation. The relation may be a general one, or one of bias, or of comparison, or of difference, or of influence, or one subject may be used as a tool for studying the other subject. This is called "Inter-Subject Phase Relation". It gives rise to a Complex Subject.

Example:

Relation of Mathematics to Astronomy
Mathematics biased to Engineering
Mathematics compared with Statistics
Application of Statistics to the study of the variation in the colour of skin in different racial groups
Influence of Physical Sciences on Biology

4 Loose Assemblage 2.- In this mode, two or more isolates from one and the same schedule are brought into mutual relation. The relation may be any one of the kinds mentioned under Loose Assemblage 1. This is called "Intra-Schedule Phase Relation". It gives rise to "Complex Isolate of Kind 1".

Example:

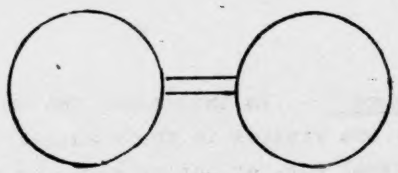
Sociology (BS). Forest Dwellers (I). Colour of skin influenced by Habitat (Complex Isolate of Kind 1).

5 Loose Assemblage 3.- In this mode, two or more isolates taken from one and the same array of order higher than 1 in one and the same schedule are brought into mutual relation. This is called "Intra-array Phase relation". It gives rise to "Complex Isolate of Kind 2".

Example:

Sociology (BS). City resident compared with Rural resident (Complex Isolate of Kind 2).

DIAGRAMMATIC REPRESENTATION OF LOOSE ASSEMBLAGE IN GENERAL



6 Fission.- In this mode, a Basic Subject or an isolate is split or fissioned into sub-divisions.

61 Fission of Basic Subject.- The non-Main Basic Subjects, formed by fission of a Main Basic Subject, may be one of four varieties:

1 Canonical Basic Subject;
2 Specials Basic Subject;
3 Environmental Basic Subject; and
4 Systems Basic Subject.

225[illegible]

Example:

11 Canonical Basic Subject:

Fissioning of the Main Basic Subject "Physics" into

- Properties of matter
- Sound
- Heat
- Radiation
- Electricity
- Magnetism

12 Fissioning of the Main Basic Subject "Philosophy" into

- Logic
- Epistemology
- Metaphysics
- Ethics
- Aesthetics

2 Specials Basic Subject

21 Fissioning of the Main Basic Subject "Medicine" into

- Child medicine
- Old age medicine
- Female medicine

22 Fissioning of the Main Basic Subject "Economics" into

- Private enterprise
- Public enterprise
- International enterprise

3 Environmental Basic Subject

31 Fissioning of the Main Basic Subject "Botany" into

- Desert botany
- Tropical botany

32 Fissioning of the Main Basic Subject "Economics" into

- War economics
- Colonial economics

4 Systems Basic Subject

41 Fissioning of the Canonical Basic Subject "Geometry" into

- Non-Euclidian geometry
- Elliptic geometry
- Hyperbolic geometry

42 Fissioning of the Main Basic Subject "Psychology" into

Experimental psychology
Psychoanalysis.
Gestalt psychology
Field psychology

62 **Fission of Isolate**.- The sharpening of the isolate resulting from fission may be represented in one of two ways:

1 As an Array division, just as in the case of fissioning of Basi Subject mentioned above; or

2 As a combination of a Principal isolate and a Speciator.

Example:

1 Fissioning of the isolate "Credit" in Economics represented by the following Array division:

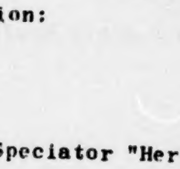
Loan
Overdraft
Assistance

2 Combination of the Speciator "Hercules Brand" with the Principal isolate "Bicycle". Thus, "Bicycle-Hercules Brand".

Dissection:- This term is used to denote fission when we consid the array of divisions of an isolate or of a Basic Subject resulting from fission.

Denudation:- This term is used to denote fission when we consider one and only one subdivision of an isolate or of a Basic Subjec resulting from fission.

DIGRAMMATIC REPRESENTATION OF FISSION IN GENERAL



7 **Fusion**.- In this mode, two or more Main Subjects are fused together in such a way that each of them loses its individuality with respect to the schedule of isolates needed to form the compound subje going with it. It gives rise to a Fused Main Subject.

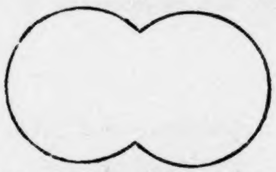
Example:

1 Fusion of the Main Basic Subjects "Biology" and "Chemistry", giving rise to the Fused Main Subject "Biochemistry".

2 Fusion of the Main Basic Subjects "Statistics" and "Economics giving rise to the Fused Main Subject "Econometrics".

[The fusion of the Main Basic Subject "Statistics" with different Main Basic Subjects has given rise to a number of Fused Main Subjects. Such a Fused Main Subject is placed as an Adjunct Main Subject to the fusing Main Basic Subject other than Statistics. A number of similar Fused Main Subjects have been recognised arising from the fusion of the Main Basic Subject "Cybernetics" and different Main Basic Subjects.] (35).

DIAGRAMMATIC REPRESENTATION OF THE PROCESS OF FUSION IN GENERAL



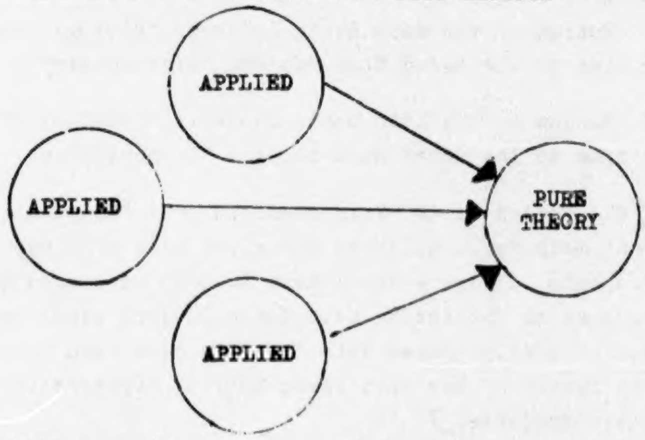
8 Distillation.- In this mode, a pure discipline is evolved as a Main Subject from its appearance-in-action in diverse compound subjects going with either different Basic Subjects or one and the same Basic Subject. It gives rise to Distilled Main Subjects.

Example:

Management Science	Evaluation technique
Experiment technique	Pure discipline of Archaeology

[The position of a new Distilled Main Subject in the schedule of Main Subjects may depend upon whether its appearance-in-action has a host Basic Subject in the Natural Sciences alone, or in the Social Sciences alone, or in the Natural Sciences as well as the Social Sciences].

DIAGRAMMATIC REPRESENTATION OF DISTILLATION



9 Partial Comprehension.- Several Main Subjects are sometimes treated integrally or disjunctively in one and the same book. A partial comprehension of this kind has meaning only with reference to the Main Subjects recognised and enumerated in the schedule of a scheme for classification.

Example:

Natural Sciences
Physical Sciences
Biological Sciences
Humanities

Language and Literature
Social Sciences
History and Political Science
History and Law

10 Subject Bundle.- For convenience in organising research, the preliminary findings and data which are obtained in different subjects going with different Basic Subjects, involved in the study of some phenomenon or entity, are published in one and the same book disjunctively for further investigation and independent elaboration by the specialists in the respective subjects. Thus, in this kind of partial comprehension, accounts of different subjects are brought together out of exigency without any substantial integral account of them.

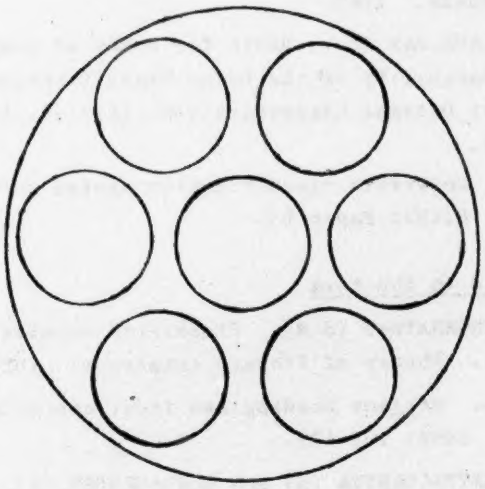
OSS 229

Example:

Ocean sciences.
Space sciences.
Surface sciences.
Defence sciences.

⎡Partial Comprehension and Subject Bundle are largely the result of developments in book production rather than being intrinsic to subjects⎦.

DIAGRAMMATIC REPRESENTATION OF PARTIAL COMPREHENSION AND SUBJECT BUNDLE IN GENERAL



BIBLIOGRAPHICAL REFERENCES

A General Reading

1 BRITISH STANDARDS INSTITUTION. Guide to the Universal Decimal Classification (UDC). (BS 1000C: 1963).

2 GOPINATH (M A). Colon Classification in the seventies. (Under publication).

3 MILLS (J). Universal Decimal Classification. 1964.

4 NEELAMEGHAN (A). Classification, Theory of (*In* Lancour (H) and others. *Ed*. Encyclopedia of library and information science. 1971. V5).

5 RANGANATHAN (S R). Descriptive account of Colon Classification. Indian ed. 1967. (US ed entitled Colon Classification. 1965).

6 --. Prolegomena to library classification. Ed 3. Assis by M A Gopinath. 1967.

7 -- and NEELAMEGHAN (A). Basis for study of compatibility and compatibility of the Colon Classification with the Universal Decimal Classification. (Lib sc. 7;1970; Paper E).

8 -- and --. Universal classification system and FID. (Lib sc. 8;1971;Paper D).

B Documents cited in the text

1 Sec 12 RANGANATHAN (S R). Classified catalogue code. 1934.

Sec 13 --. Theory of library catalogue. 1938.

2 Sec 13 --. Subject heading and facet analysis. (J doc. 20; 1964; 109-19).

3 Sec 13 BHATTACHARYYA (G) and NEELAMEGHAN (A). Postulate-based subject heading for dictionary catalogue system. (Annual seminar (DRTC). 7;1969; Paper CA).

Sec 13 NEELAMEGHAN (A) and BHATTACHARYYA (G). Chain procedure and micro subjects. (Lib sc. 5;1968; Paper E).

	Sec 13	RANGANATHAN (S R). Classified catalogue code etc. Ed 5. 1964. Part K.
	Sec 13	--. Subject heading and document finding system. (Annual seminar (DRTC). 7;1969; Paper CB).
4	Sec 14	NEELAMEGHAN (A). Analysis of readers' questions: A case study. (Lib sc. 6;1969; Paper N).
5	Sec 151	--. Integrated approach of India to the design and developments of documents retrieval systems. (International forum on informatics (Moscow). Volume of papers. 1969. V2, Sec 12-16.)
6	Sec 152	MILLER (G A). Language and communication. 1961. P 211.
7	Sec 153	NEELAMEGHAN (A). Integrated approach of India to the design and development of document retrieval systems. (International forum on informatics (Moscow). Volume of papers. 1969. V2, Sec 3-5).
8	Sec 16	RECORDING THE subject interest of subject specialist. (Seminar on reference service (DRTC). Working paper. 1971. Chap BE).
		McCASH (W H) and CARMICHAEL (JJ). UDC user profile as developed for a computer-based SDI service in the iron and steel industry. (J doc. 26;1970; 295-312).
9	Sec 2	NEELAMEGHAN (A). Classificationist and the study of the structure and development of the universe of subjects. 1967. (FID/CR report series. 5).
		--. Research on the structure and development of the universe of subjects. (Lib sc. 4;1967; Paper Q).
10	Sec 21	RANGANATHAN (S R). Basic subjects and their kinds. (Lib sc. 5;1968; Paper C).
		--. Colon classification, Ed 7 (1971): A preview. (Lib sc. 6;1970; Paper M).
		--. Compound isolate and compound basic subject: Evolution of the concept through forty years. (Lib sc. 7;1970; Paper A).

		--. Prolegomena to library classification. Ed 3. 1967. Part P.
11	Sec 21	DEVADASON (F J). Subject specialist's view of the universe of subjects: Some examples. (Annual seminar, (DRTC). 7;1969; Paper AA).
12	Sec 221	BAVADEKAR (P N) and others. Rate of development of the universe of subjects. (Annual seminar, (DRTC). 5;1967; Paper A).
13	Sec 221 Table 5	CARTER (L F) and others. National document-handling systems for science and technology. 1967. P 302.
14	Sec 222	PRICE (D J De S). Little science, big science. 1963. P 24.
15	Sec 222 Sec 81	NEELAMEGHAN (A). Comparison of "Subject" and "System". (Annual seminar,(DRTC). 8;1970; Paper BB).
16	Sec 223	RANGANATHAN (S R). Descriptive account of Colon classification. Indian ed. 1967. Chap V.
17	Sec 223	--. Kinds of bonds between two subjects including fusion bond. (Annual seminar, (DRTC). 5;1967; Paper B).
18	Sec 231	JAYARAJAN (P). New basic subjects, partial comprehensions and subject bundles. (Annual seminar, (DRTC). 8;1970; Paper BC). RANGANATHAN (S R). Colon classification, Edition 7. (1971): A preview. (Lib sc. 6;1969; Paper M).
19	Sec 232 Table 8	GOPINATH (M A), and JAYARAJAN (P). Analysis of subjects: A case study. (Annual seminar, (DRTC). 7;1969; Paper EA).
20	Sec 31 821	--. Hidden roots of classification. (Lib sc. 4;1967; Paper A, Sec 84-85).
21		--. Doc-finder. (Lib sc. 5;1968; Paper N, Sec 52).
21	Sec 32	HUTTON (G H). Product analysis by coordinate index. (Aslib proc. 20;1968; 172).
22	Sec 52	RANGANATHAN (S R). Prolegomena to library classification. Ed 3. 1967. Part Q.

233

23 Sec 61 RICHARDSON (E C). Classification. 1912.

24 Sec 61 SAYERS (W C B). Introduction to library classification. 1918.

25 Sec 61 BLISS (H E). Organisation of knowledge in libraries. 1933.

26 Sec 62 RANGANATHAN (S R). Five laws of library science. 1931. Ed 2. 1957. Ed 3 (in preparation).

27 Sec 72 DONKER DUYVIS (F). Policy of revision of the Universal Decimal Classification (1). (Rev doc. 23;1956; 140).

PERREAULT (J M). Towards a theory for UDC. 1969.

28 Sec 73 RANGANATHAN (S R). Prolegomena to library classification. Ed 3. 1967. Chap SB.

29 Sec 822 KILPATRICK (W H). Philosophy of education. 1951. New York, Macmillan. [Also published in Park (Joe), Ed. Selected readings in the philosophy of education. Ed 3. 1968. P 107-17].

30 Sec 822 RUSSELL (Bertrand). Has man a future. 1961. (Penguin special). P 8.

31 Sec 825 RANGANATHAN (S R). Prolegomena to library classification. Ed 3. Assis by M A Gopinath. 1967. Part F.

32 Sec 825 --. Elements of library classification. Ed 3. 1962. Sec N32.

33 Sec 91 BHATTACHARYYA (G). Vital role of depth classification in a system for document finding: A trend report. (Lib sc. 6;1969; Paper C).

34 Sec 92 GUPTA (B S S). Program-package for a system for document finding. (Lib sc. 7;1970; Paper H). [This paper refers to the other papers reporting work done in DRTC on a computer-aided system for document finding].

35 Sec 94 FOSKETT (D J). Some fundamental aspects of classification as a tool in informatics. (In Mikhailov (A I) and others, Ed. On theoretical problems of informatics. 1969. P 70).

36 Sec 94 LANGRIDGE (D W). [Review of] Foskett (A C): Subject approach to information. 1969. (Library Assoc rec. 72;1970; P 344).

37 Sec 951 RANGANATHAN (S R) and NEELAMEGHAN (A). Universal classification system and FID: A note on FID/CR report (9). (Distributed to members of FID/CR and FID/CCC. 1970).

38 Sec 96 DONKER DUYVIS (F). Jubilee of the creator of the Colon Classification. (In Kaula (P N), Ed. Library science today. 1965. Chap B1).

39 Appendix 2 Sec 7 GOPINATH (M A). Reception of cybernetics in Colon classification. (Lib sc. 7;1970; Paper K).

SOME NOTATIONAL,
HIERARCHIC AND SYNTACTIC PROBLEMS IN CONNECTION WITH CONCORDANCES BETWEEN UDC AND THESAURI

Einar ÖHMAN, Christina OLIVECRONA
Swedish Society for Technical Documentation Ingeniörsvetenskapsakademien, Biomedical Documentation Center, Stockholm, Sverige

1. Introduction

In the autumn of 1970 a working group appointed by the FID Central Classification Committee, FID/CCC, started a pilot project on a concordance between UDC and the Thesaurus of Engineering and Scientific Terms (TEST) (1). A first progress report, elaborated by the coordinator of the working group, Mr. H. Wellisch, was presented to FID/CCC at their meeting 10-12 May 19[illegible] (2).

At a fairly early stage of the work it became evident that it should be possible to convert a very high percentage of TEST terms into exactly corresponding UDC codes. This fact gave rise to the thought that the concordance problem might be profitably studied also along other lines than those directly connected with the TEST-UDC project. One of the present authors (Öhman), member of the FID/CCC working group just mentioned, therefore left it to his colleagues, Messrs. I. de Vries and H. Wellisch, to further elaborate the progress report to FID/CCC into a paper suitable for the Herceg Novi symposium (3), and joined forces with the other of the present authors (Olivecrona).

It was considered of interest to find out, if terms from thesauri in other fields of knowledge could be converted into UDC codes as satisfactorily as the Wellisch report indicated with regard to TEST. A similar concordance study was thus carried out between the MEDLARS thesaurus, Medical Subject Headings (MeSH) (4) and UDC. A brief study was also made of the concordance prepared between UDC and part of Thesaurus Documentaire d'Economie by a working group under FID/C3, the international UDC committee for social sciences (5,6).

The results reported by Wellisch were considerably more favourable than those found in 1969 by an ASLIB working group (7). They were only able to convert 52 per cent of a big sample of TEST terms into UDC codes, while the figure reported by Wellisch for a different sample was in the neighbourhood of 90 percent. We have made an attempt to elucidate the reasons for this discrepancy.

Considerable space has been devoted to problems with hierarchies in UDC in relation to TEST and MeSH. The fact that MeSH in addition to its main sequence of terms, the "main headings", also has a series of "subheadings", which are applied to the main headings according to fixed rules, made it natural to compare the function of these subheadings with the extensive series of auxiliaries in UDC class 61, medicine. It is hoped that this study shall prove useful in connection with the current revision of UDC 61.

The studies reported in this paper on the ASLIB investigation of classification in science information, on the pilot concordance between Thesaurus documentaire d'Economie and UDC, and on some hierarchical problems in TEST, were made by E. Öhman. The sections dealing with comparative studies between MeSH and UDC with regard to hierarchy and syntax were written by C. Olivecrona. The pilot concordance between MeSH and UDC was elaboratated in close cooperation by both authors, who are also jointly responsible for all conclusions made.

The following UDC editions and supplements have been consulted in connection with the present study:

Full edition

Universal Decimal Classification. Full English Edition.
  Fourth International Edition. London.
    The numerous parts of this edition published 1968-1970 have been especially useful.

DK Dezimalklassifikation. Deutsche Gesamtausgabe. Dritte und achte
  Internationale Ausgabe, Berlin.
    This edition is kept up to date by means of supplementary sheets published in DK-Mitteilungen.

Medium edition

DK-Handausgabe, I & II, Berlin 1967/1968.

Abridged edition

Universal Decimal Classification. Abridged English Edition. 3rd edition, London 1961.

10-Year Supplement to Abridged UDC Editions 1958-1969 (French - English - German). The Hague 1969.

Supplements

Cumulated UDC Supplement (-1964), Vol I - VI, The Hague 1967.

Extensions and Corrections to the UDC, Series 6, No. 6. The Hague 1968.

Extensions and Corrections to the UDC, Series 7, No. 5. The Hague 1971.

Various special subject editions in the English language have been occasionally consulted with regard to English terminology.

2. Pilot concordances between thesauri and UDC

2.1. The ASLIB study on Classification in Science Information

The pilot concordance between TEST and UDC already mentioned will not be discussed in any detail in this paper, but the results obtained will be referred to in various connections. On the other hand it seemed important to deal with the ASLIB comparative study of classification systems. This was evidently not available to H. Wellisch when he wrote his report to FID/CCC on the TEST/UDC pilot study. We received a copy in the middle of April this year.

The study on Classification in Science Information (7).was carried out in 1969 by an ASLIB working group for the International Council of Scientific Unions (ICSU) and was discussed in December of that year at the fourth session of the UNESCO-ICSU joint central committee to study the feasibility of a world science information system. Part of the investigation is of primary interest to the present study. We refer to the detailed comparisons between thesauri and four conventional universal classification systems: Dewey, UDC, Bliss and Colon. Three of these classifications do not concern us, and we will thus concentrate on the comparison between TEST and UDC.

The ASLIB report is marked "Distribution : restricted". Nevertheless, some of the results reported, especially with regard to UDC, have obtained considerable publicity through an article published by H. Arntz in 1970 (8). There can be no doubt that the account given by him of the ASLIB findings, have been generally interpreted as a sharp criticism of the efficiency of UDC in its present shape as an indexing language.

The ASLIB comparison comprises no less than 1893 TEST terms taken from 54 of the about 700 hierarchies or "descriptor families" found in the last part of TEST, the Hierarchical Index. The number of terms in the hierarchies investigated ranges from 180 (Bacteria) to 5 (Stochastic processes). In accordance with the main object of the study, the hierarchies chosen were taken from science rather than technology.

Of the 1893 TEST terms only 961 or about 52 per cent could be converted into UDC notations by the ASLIB working group. Even if this figure is very much higher than those obtained for the other classification systems, it is definitely disappointing. The results just reported from the first part of a pilot study by a working group appointed by FID/CCC, point in an entirely different direction. As stated by Wellisch, only one of the 173 terms investigated was lacking in UDC. Of the remaining 172 terms, no less than 90.6 per cent could be given an exact UDC equivalent. The remaining 9.4 per cent could be expressed by subsuming them under broader terms or by constructing a code that somehow approximated the meaning of the TEST descriptor. The sample is admittedly small, but as one term was taken from each of the TEST category groups, it is believed that the results are at least roughly representative of TEST as a whole. In his conclusions Wellisch therefore considered himself justified to state "that it is possible to find corresponding UDC codes for a very large percentage (possibly up to 90 %) of all descriptors in TEST without sacrifying exactness in the meaning of terms".

It seems important to find out the reasons for the great discrepancy between the results of the two investigation. It has only been possible to make a few fairly small samples, but we believe that these are sufficient for our purpose. The ASLIB results are summarised in their Table 15, which gives the number of TEST terms and the number of UDC codes found for each of the 54 hierarchies. All instances when the highest term could not be denoted by a UDC code are marked with an asterisk. In Table 14 the full results obtained on one hierarchy, Thermodynamic properties, are given as an example.

2.1.1. The highest terms in the hierarchies

In six of the 54 hierarchies the UDC code corresponding to the highest term is indicated as not found. This is correct if we confine ourselves to main UDC numbers, but there is no reason why auxiliaries should not be used. In this way four of the terms may be denoted as follows:

Body fluids = 616-008.8
The term appears in the alphabetic index of the abridged English edition.

Chemical properties = 54.04
The term appears both in the abridged and the full English edition.

Congenital abnormalities = 616-007-053.1
The term appears in the alphabetic index of the abridged English edition. The full English 616 had not appeared when the investigation was made, but the code is easily found, e.g. in DK-Handausgabe.

Electrical properties = 537.081.7
The English full edition has the text "Specific constants for substances" for this auxiliary. It has been revised, and the current text is found in the German class 5 (1958): "Stoffwerte. Material-Konstanten".

We have not been able to find any UDC codes corresponding to the two remaining terms, Bosons, and Stochastic processes. These terms are very specific, and it is surprising to find them at the head of hierarchies. Bosons will be discussed in the following section, Stochastic processes may well find its way into UDC in connection with the revision of class 51, Mathematics, now in progress.

2.1.2. Two small hierarchies

A special study was made of two small hierarchies, i.e. those headed by the terms Bosons and Crystals. The choice was made for several reasons, time did not permit exploration of any of the larger hierarchies, we found the term Bosons rather intriguing from a hierarchical point of views, and the experience of one of the authors made him expect a considerably higher percentage of UDC codes than that found by the ASLIB group in the fields of elementary particles and crystals.

It is very surprising that UDC codes could only be found for three of the particles appearing as subdivisions of Bosons. In fact, no less than seven of the eight terms or their synonyms appear in the special subject edition for Nuclear Science and Technology (in English!) published in 1964. They also appear (in English and German) in Extensions and corrections, series 4, or Cumulated UDC supplement. This means a rise in percentage of corresponding UDC codes in this hierarchy from 33 to 79 (see Table 1).

Table 1. TEST Hierarchical Index: BOSONS

TEST term	UDC code	UDC text
Bosons	539.12Bosons	Elementary and simple particles/Bosons
Alpha particles	539.128.4	Alpha particles
Deuterons	539.128.2	Deuterons
Helium 4	546.291.027.4	Helium isotope 4
Mesons	539.126	Mesons and hyperons
	539.126.3/.4	Mesons
Kaons	539.126.4	Heavy mesons
Meson resonances	-	-
Pions	539.126.34	Pi mesons
Photons	539.122	Photons

The elementary particles are sometimes divided into two main groups according to the type of spin, Fermions and Bosons. Both appear in TEST, the former has Elementary particles as BT, the latter has no BT. In the hierarchical index we thus find a rather big hierarchy headed by the term Elementary particles and a smaller one Bosons. The former contains all the terms under Bosons which are elementary particles: mesons, kaons, meson resonances, pions and photons. In addition to these the Boson hierarchy contains the terms alpha particles, deuterons, and helium 4. This does not agree with the definition of the term Bosons available to us. Helium 4, a stable isotope of an element, has a UDC number in 546, the seven particles, whether elementary or not, are all found in 539.12. The term for which we could not find a UDC equivalent is Meson resonances.

Table 2. TEST Hierarchical Index: CRYSTALS

TEST term	UDC code	UDC text
Crystals	548.0	General science of crystals
Bicrystals	548.24	Twin crystals
Dendritic crystals	548.232.4	Dentritic stucture
Etched crystals	548.574	Etched figures
Ferroelectric crystals	548.0:537.226	Crystals: Ferroelectricit
Ionic crystals	548.312.2	Pure ion lattices
Metal crystals	548.312.6	Metal lattices
Piezoelectric crystals	548.0:537.228.1,	Crystals:Piezoelectricity
Polycrystals	548.23	Crystalline agregates
Single crystals	548.55	Single crystals
Whiskers(single crystals)	548.55 Whiskers	Single crystals/Whiskers
Metal whiskers	548.312.6:548.55Wh	Metal crystals: Single crystals/Whiskers

The next small hierarchy studied is Crystals. The number of terms is 12, not 13, as stated in table 15, as one term appears twice. UDC equivalents were found by the ASLIB group for six terms or 50 per cent. It seems safe to assume that these are the six terms that appear in the full English UDC edition (1943). Of the remaining terms, two correspond exactly to UDC numbers in the full German edition (1958), ionic crystals and metal crystals, and two may be easily expressed by means of colon, ferroelectric crystals and piezoelectric crystals. The two remaining terms have no exact UDC equivalents, Whiskers (single crystals) and Metal whiskers. One has to be content with using the number for single crystals in the first case and a colon combination to this number in the second. In any case, 10 of the twelve terms have exact UDC equivalents, which means 83 per cent (see table 2).

2.1.3. Thermodynamic properties

As already mentioned, table 14 of the ASLIB working group report contains the full comparison between one TEST hierarchy and UDC. The hierarchy is headed by the term Thermodynamic properties. The UDC number assigned to this term is 536.7, preceded by the sign < "less than", indicating that the compass of the term is narrower than that of the UDC code. A study of the 28 UDC codes assigned by ASLIB to subordinate terms, shows, that only four are subdivisions of 536.7. The remainder is spread over the whole of 536. In fact, one number is a subdivision of the neighbouring group 537. An explanation of this curious fact can be found from the following entry in TEST: "Thermal properties USE Thermodynamic properties". The first edition of the EJC thesaurus (9) still contains the term Thermal properties, and it seems rather doubtful if this change of terminology has been universally accepted. This is an example of the danger of comparing terms from different indexing systems without carefully considering their meaning. In TEST the two terms Heat and Thermodynamics stand isolated with neither BT nor NT but both with a number of RT's (Related terms). A comparison between the category group Thermodynamics (which contains the term Heat) and the hierarchy Thermodynamical properties, shows, that the term Thermodynamics could well be denoted 536 (text in the English UDC editions Heat. Thermodynamics), and that Thermodynamic properties could be rendered 536.081.7.

The ASLIB group could not find any UDC codes for the intermediate headings in the hierarchy, Thermochemical properties and Thermophysical properties. This is surprising with regard to the former term, as all its 18 subdivisions are found in 536.65 or 536.66. The obvious UDC code is thus 536.65/.66.081.7. The term Thermophysical properties undoubtedly causes some difficulties. The 20 subdivisions of this term are scattered in UDC. Thus five terms are found in other parts of class 53, one in 532, one in 535 and three in 537. Of the remaining fifteen, three are found in 536.2, ten in 536.4, one in 536.6 and one in 536.7. If we study the TEST category groups instead of the hierarchy we find a similar scattering of thermophysical terms:

3 terms in category group 0704 Physical and general chemistry
2 " " " " 1407 General concepts
3 " " " " 2003 Electricity and magnetism
1 " " " " 2012 Solid state physics
11 " " " " 2013 Thermodynamics

The terms most naturally denoted by UDC numbers outside 536 might, if seen specifically from the point of view of heat, be coloned to subdivisions of 536, thus 536.2..:532.., 536.3..:535.., 536.5..:537.. . We have therefore not hesitated assigning 536.2/.5.081.7 to the term Thermophysical properties.

Table 3. TEST Hierarchical Index: Thermodynamic properties.

TEST term	UDC code ASLIB	UDC code Ö. & O.	UDC text Ö. & O.
Thermodynamic properties	536.7	536.081.7	Thermal properties
Dissociation energy	–	536.657	Heat of dissociation
Enthalpy	–	536.722	Enthalpy
Entropy	536.75	536.75	Entropy
Free energy	536.77	536.77	Free energy
Gibbs free energy	536.775	536.775	Gibbs-Helmholtz principle
Helmholtz free energy	536.775	536.775	" " "
Solidification points	536.421.4	536.421.4	Freezing point
Thermal expansion	536.41	536.41	Thermal dilatation
Thermochemical properties	–	536.65/.66	Latent heats and heats of reaction
Heat of ablation	–	–	–
Heat of absorption	–	536.65:542.744	Heat of absorption
Heat of activation	–	536.65:541.124.7	Heat of activation
Heat of adsorption	536.658	536.558	Heat of adsorption
Heat of crystallization	–	538.65:532.78	Heat of crystallization
Heat of mixing	–	536.65:542.63	Heat of mixing
Heat of reaction	536.66	536.66	Heat of reaction
Calorific value	–	536.62.081.7	Calorific value
Heat of combustion	536.662	536.662	Heat of combustion
Heat of dissociation	536.657	536.657	Heat of dissociation
Heat of formation	–	536.66	Heat of formation
Heat of hydration	536.664	536.664	Heat of hydration
Heat of solution	536.653	536.653	Heat of solution
Heat of transformation	536.656	536.656	Heat of transformation
Latent heat	536.65	536.65	Latent heat
Heat of fusion	536.652	536.652	Heat of fusion
Heat of sublimation	536.654	536.654	Heat of sublimation
Heat of vaporisation	536.655	536.655	Heat of vaporisation
Thermophysical properties	–	536.2/.5.081.7	Thermophysical properties
Boiling point	536.423.1	533.423.1	Boiling point
Critical point	536.44	536.44	Critical points
Critical pressure	–	536.44:533.12	Critical pressure
Critical temperature	–	536.44:536.5	Critical temperature
Critical volume	–	536.44:533.11	Critical volume
Emissivity	–	535.231.2	Emissivity
Melting point	536.421.1	536.421.1	Melting point
Peltier effect	537.322	537.322.15	Peltier effect
Pyroelectricity	537.227	537.227	Pyroelectricity
Seebeck effect	–	537.322.11	Seebeck effect
Sensible heat	–	536.722	Enthalpy
Softening points	536.421.2	536.421.2	Softening point
Specific heat	536.63	536.63	Specific heat
Sublimation pressure	536.422.15	536.422.15	Vapour pressure of solids
Surface energy	–	532.614.2	Surface energy
Thermal conductivity	536.2.022	536.2.022	Thermal conductivity
Thermal diffusivity	–	536.2.023	Temperaturleitfähigkeit
Thermal stability	–	541.428.2	Substances with low reactivity
Vapour pressure	536.423.15	536.423.15	Vapour pressure
Volatility	536.423.16	536.423.16	Volatility

Of the 49 terms in the hierarchy for Thermodynamic properties, the ASLIB group succeeded in finding UDC codes for 29 terms, or 59 per cent. Of these codes 28 are single UDC numbers. Of the remaining 20 we succeeded in coding 19. There was only one term for which no UDC code could be assigned, in three other cases the meaning of the code was slightly more general than the term. The number of specific codes is thus 45, which gives a percentage as high as 92 (se table 3). Two of the terms that could not be given specific codes are Gibbs free energy and Helmholtz free energy. UDC gives Gibbs-Helmholtz principle, 536.775, a subdivision of Free energy, 536.77. We are not certain that there is literary warrant for subdividing 536.775, and are rather inclined to consider this number as a specific code for both terms. If this is correct, the percentage rises to 96.

Of the 19 terms left out by ASLIB which we could give UDC codes, 11 may be obtained from printed English schedules. One of these, Surface energy, does in fact occur specifically in the full English edition. Of the remaining 10, three require the use of a special auxiliary. In seven cases two numbers must be coloned. Eight terms require the use of a German edition. Six of these codes are single numbers, one consists of a main number with special auxiliary, and one of two coloned numbers. The term that we were not able to codify is Heat of ablation.+)

2.1.4. Summary of UDC concordance studies on three TEST hierarchies.

The three TEST hierarchies studied by us contain together 70 terms. The ASLIB group found UDC codes for 38 of these or 54 per cent. This is slightly higher than the average obtained for the 1893 terms from 54 hierarchies. We were able to find explicite codes for 62 terms or 89 per cent, in good agreement with the result reported by Wellisch. Several factors seem to contribute to the discrepancy between the two studies. The ASLIB group appears to have used coordinated codes (main number with auxiliary or coloned numbers) only very sparingly. Changed texts and new numbers appearing in Extensions and Corrections to the UDC as well as in the German class 5 and its supplements have not been used to any considerable extent (at least not in connection with the hierarchies studied by us). Occasionally, the group has also

---

+) In TEST ablation refers both to glaciers and to spacecraft.

overlooked texts in the full English edition exactly corresponding to TEST terms or synonyms to such terms. Difficult to understand are the results obtained in the hierarchy Bosons, where almost all terms or their synonyms are found in the special subject edition for Nuclear Science and Technology.

2.2. The Földi-Sommer concordance project

The Thesaurus documentaire d'Economie appeared 1969-1971 in a provisional edition of 18 sections with French terms, alphabetically arranged, and with English translations of these terms. A nineteenth section contains an English alphabetic index. Commissioned by FID/C3, the international UDC committee for social sciences, and economically supported by FID, T. Földi and K. Sommer have worked out a German language concordance between parts of this thesaurus and UDC, which was presented to FID/C3 in October 1970. Their work comprises four sections of the thesaurus with altogether 345 terms, as follows:

8. Histoire économique	15 terms
9. Doctrines économique	57 terms
14. Fluctuations et croisance économiques	27 terms
17. Facteurs de production	246 terms

Of the 345 terms 22 or 6.4 per cent were not provided with UDC codes by Messrs. Földi and Sommer. In three cases this is stated to be due to ambiguous thesaurus terms and in two others to the fact that the terms did not belong strictly to the subject matter of the section in question. The investigators remark that the English translations often do not correspond exactly to the French terms, and that the latter are often so general that a correct translation to German technical terms is difficult. In three cases no German term has been given (and of course no UDC code). In view of the fact that the thesaurus is stated to be provisional and that criticism is welcomed, the shortcomings are hardly surprising.

The concordance work aims at facilitating the revision work of FID/C3. The investigators have therefore not hesitated in substituting new UDC codes (at present inofficial) for existing codes in no less than 29 cases (8.3 %) and even assigned such codes to eight terms to which they could not find a satisfactory existing code. In a considerable number of cases they have also assigned double codes in fields where revision is in progress: one code recently cancelled or about to be cancelled, and another one recently introduced or about to be introduced.

As the concordance in its present state is thus a working paper of FID/C3, and as the thesaurus is provisional, any detailed discussion of the results obtained is undoubtedly out of place. However, some general conclusions may, we think, be safely drawn, and these could be compared with the results obtained in the technical and medical fields with other thesauri.

As may be easily ascertained, Messrs. Földi and Sommer were able to assign UDC codes, using existing UDC numbers, to 91.3 of the thesaurus terms.

To 323 thesaurus terms 368 UDC codes were assigned (not counting cancelled numbers and such about to be cancelled, which had been provided with substitutes). Editions necessary to obtain these codes and types of codes used are found in Table 4.

The number of codes requiring the medium or full edition had to be estimated, as a considerable number of the codes are so new that they do not appear in any printed medium edition.

Table 4. UDC editions and types of UDC codes in Földi-Sommer pilot concordance.

Edition	UDC code N	UDC code %
Abridged	202	54.9
Medium	70	19.0
Abridged + Medium	272	73.9
Full	96	26.1
	N	%
Simple main numbers	247	67.1
Main number + auxiliary	87	23.5
Coloned main numbers	34	9.4

2.3. Medical Subject Headings and UDC. A pilot concordance study

2.3.1. Short description of Medical Subject Headings

The MEDLARS thesaurus, Medical Subject Headings (MeSH), contains about 8000 descriptors or main headings. These are arranged in an alphabetical section with cross references of four commonly used types: See, See under, See also specific and See also related. The number of cross references is about the same as that of the descriptors. There is also a series of Categorized Lists, table 10, where the descriptors are displayed in faceted groupings, constituting 14 Categories (A - N), eight of which are divided into a total of 65 Subcategories.

For computerized retrieval purposes the hierarchical relations between the terms have been worked out more in detail, the so called Tree Structures in MeSH. Due to technical limitations, the depth of the hierarchies is limited to four levels. The same descriptor can appear at the most in four different positions in the tree structures. These structures differ from a strictly hierarchical classification in as much as the thesaurus principle of alphabetical order is kept in each level.

Apart from the tree structure, which constitutes the paradigmatic structure in MeSH, there is superimposed a system of subheadings, which may be regarded as a precision device of syntactic nature. Each of the 60 subheadings is provided with a detailed definition, and it is clearly indicated in which categories it may be applied.

The principles of coordinate indexing in MEDLARS relies on the following three coordinative devices:
a. Precoordination in main headings, e.g. Neoplasm transplantation
b. Postcoordination of main headings, e.g. Tibia and Elasticity
c. Main heading / subheading combination, e.g. Glucose / metabolism

2.3.2. The MeSH / UDC pilot concordance

A comparative study of MeSH and UDC must take into account not only the number of corresponding terms and concepts, but also the different UDC structures in their relation to the paradigmatic as well as to the syntactic structures in MeSH. This will be done in later sections of this paper. Here we will confine ourselves to an account of a simple pilot concordance between MeSH main headings and UDC codes, similar to that reported by Wellisch for TEST and UDC. For the sake of uniformity the same sampling method was used as in the last mentioned study, but in order to obtain a sufficiently big sample, we took two terms from each of the 65 subcategories as well as from each of the six categories not provided with subcategories. The sample thus consisted of 142 terms. The first term in each pair had the same initial letter as the title of the subcategory. The second term simply was the one following directly after the first. The capacity of UDC to code these concepts was studied, full use being made of all devices applicable, auxiliaries, colon and the apostrophe.

It was sometimes possible to assign more than one UDC code, e.g. one taken from class 61, medicine, and another from 57, biology or 54, chemistry. The MeSH subcategories and descriptors are so clearly defined, however, that we have found it possible always to present only one UDC code for each descriptor. The results of the pilot study are summarized in tables 5-9.

Table 5. UDC codes required for 141 of 142 terms from 71 MeSH categories

UDC codes consisting of	N	%
Single main number	72	51.1
Coloned main numbers	4	2.8
Single main number with auxiliary	50	35.5
Coloned main numbers with auxiliary	13	9.2
Single main number with apostrophe	2	1.4
Coloned numbers, with or without auxiliaries	17	12.0
Numbers with auxiliaries, with or withour colon	63	44.7
Coordinated numbers (aux., colon, & apostrophe)	69	48.9

From table 5 can be seen that it was possible to assign UDC codes to 141 of the 142 terms. Of these codes 72, or 51.1 per cent, consist of simple main UDC numbers, while 69, or 48.9 per cent are coordinated. In the latter group of codes the simple main numbers were modified by one or more auxiliaries, by colon combination with another main number, by apostrophe, or by a combination of colon and auxiliaries. Auxiliaries appear in altogether 63 codes (44.7 %) and colon combination in 17 codes (12.0 %). The apostrophe device was only used twice. The term for which we could find a UDC equivalent is Intrinsic factor.

The MeSH categories may be roughly diveded into two groups, one specialised and strictly medical (categories A, C, D, E, and F), the other containing subject matter of a more general nature (categories B and G - N). In the first group 88 per cent of 103 codes contain a UDC number from class 61, medicine, in the second group only 16 per cent of 38 codes.

Table 6. Simple UDC main numbers in different concordances

Thesaurus	Simple main numbers
TEST	68.7 %
TDd'Ec	67.1 "
MeSH A - N	51.1 "
" A, C - F	46.5 "
" B, G - N	63.2 "

Table 6 shows, that while about two thirds of the codes in the samples from TEST and the economic thesaurus consist of simple UDC numbers, only about on half of the codes in the Mesh sample are of this type. A still lower figure was obtained for the "special" MeSH sample, while that for the "general" MeSH sample (unfortunately too small to be statistically significant) approaches those for the other thesauri.

Table 7. Exact and approximate UDC equivalents of 142 terms from 71 MeSH categories

Categories	Terms	Exact equiv.		Approx. equ.		No equiv.	
		N	%	N	%	N	%
A - N	142	106	74.7	35	24.6	1	0.7
A, C-F (88 % UDC 61)	104	74	71.2	29	27.8	1	1.0
B, G-N (16 % UDC 61)	38	32	84.2	6	19.8		

The same tendency is evident in table 7, where the percentage of exact UDC equivalents to MeSH descriptors is shown. In comparison to the result of the TEST pilot study, about 90 per cent, the figure obtained, 74.7 per cent, is surprisingly low. A still lower figure was found for the "special" group of categories. We refer here to exactly equivalent meaning, not to identity of terms.

Table 8. UDC editions required for coding 141 of 142 terms from 71 MeSH categories

Categories	Terms	Full ed.		Medium ed.		Abridged ed.	
		N	%	N	%	N	%
A – N	141	60	42.6	40	28.4	41	29.0
A, C–F (88 % UDC 61)	103	55	53.4	31	30.1	17	16.5
B, G–N (16 % UDC 61)	38	5	13.4	9	23.6	24	63.0

Table 9. Use of different UDC editions in thesaurus-UDC concordane

Edition	TEST	TDd'Ec	MeSH		
			A – N	A, C–F	B, G–N
	%	%	%	%	%
A	61.9	54.9	29.0	16.5	63.0
M	18.7	19.0	28.4	30.1	23.6
A + M	80.6	73.9	57.4	46.6	86.6
F	19.4	26.1	42.6	53.4	13.4

The most spectacular difference between the results from the MeSH sample and those from the other thesauri, however, is shown in tables 8 and 9. Wellisch stated in his first report to FID/CCC on the TEST pilot study, that only about 20 per cent of the TEST descriptors would require the full UDC edition for expressing an adequate UDC code. According to the pilot study on the economic thesaurus, the corresponding figure in that field is about 26 per cent. As shown in table 8, the figure for the MeSH sample is about twice as high, 42.6 per cent, and for the "special" MeSH categories still higher, 53.4 per cent. It is evident, that the selection of concepts and corresponding UDC numbers given in the abridged, and even in the medium edition, is completely inadequate for the detailed indexing of medical literature. An obvious explanation of this fact is given by K. Fill in his introduction to UDC (10). In table on p. 48 he shows the total sum of UDC numbers devoted to various parts of the UDC in the German full and medium edition. The medium edition percentage is considerably lower for class 61 than for any other part of the UDC as seen from the following extract:

UDC	Full edition	Medium edition	% of full ed.
3	14 695	4 534	31
5/539	9 228	2 319	25
54/59	15 143	4 249	28
6/61	13 620	2 253	17
62/621	13 440	3 185	24
66	12 454	3 806	30
Total	130 184	36 753	28

A few words more ought to be said about the approximate equivalents mentioned above. We have interpreted the word "exact equivalent" very strictly, and some of the codes counted as "approximate" have a meaning that comes very close to that of the corresponding descriptor.Generalyy these UDC codes are somewhat wider in compass than the descriptors, but in two cases we came closer to the descriptor meaning by choosing a UDC code of slightly narrower compass. In one case the meaning of the UDC code is rather "related" than narrower or broader than that of the MeSH term. In two cases we have taken the liberty of extending 611, anatomy, in parallel with 616, thus we have used 611.314.13 for enamel and 611.314.16 for root and cement of the teeth from the anatomical point of view.

A few examples of "approximate" codings are given below:

Descriptor	UDC code	UDC text	Note
Sacrococcygeal region	611.99	Coccygeal region	UDC narrower
Trophoblast	611-013.31	Segmentation, morula, blastula	UDC broader than MeSH
Mandragora	582.951.4	Solanacae	UDC broader than MeSH
Narcolepsy	616.8-009.836.12	Somnolence, drowsyness, narcolepsy	UDC broader than MeSH

In most of these cases it is possible to obtain an exact equivalent by resorting to the A/Z device.

The full tables of the MeSH / UDC concordence study will be deposited at the FID Classification department, The Hague, and will also be available at Biomedical Documentation Center, Karolinska Institutet, Stockholm.

2.4. Some conclusions from the concordance studies

The discussion presented in this paper on some parts of the ASLIB study of Classification in Science Information should, we hope, tend to increase the confidence in the results reported by Wellisch on the work of the FID/CCC working group UDC/TEST. We don't want to anticipate the results from the further work of this working group, but from the knowledge we have of the material collected, we have no reason to expect a change in the overall picture in an unfavourable direction. We consider it well established that a very high percentage of TEST descriptors can be adequately translated into UDC codes. Continued work on these lines could undoubtedly be utilized to great advantage in connection with UDC revision work.

The concordance work in the economic field is done in close cooperation with the international UDC committe FID/C3, social sciences. It is evident that this work could be of great benefit both to the UDC revision work and to the further development of the thesaurus.

The position is rather different in the medical field. The fact that there are about twice as many UDC numbers in class 61 as there are purely medical descriptors in MeSH, taken in connection with the result just reported, that only 75 per cent of the descriptors could be given exactly corresponding UDC codes, seem to indicate that some other factors, not very evident in the other concordance studies, are involved.

3. Hierarchical studies

This section will be devoted to a study of some hierarchical aspects of the TEST and MeSH thesauri. There is undoubtedly a big difference between these two thesauri with regard to hierarchical structure. TEST is a representative of a common type of thesaurus, where hierarchy plays a very minor part, while MeSH on the other hand has a much more complicated structure with important hierarchical features.

3.1. Some notes on the TEST hierarchies

The hierarchies in TEST have already been touched upon to some extent in the section dealing with the ASLIB study on Classification in Science Information. They shall be further elucidated in this section, where a few examples will be analysed.

If you look up the term Pyroelectricity in TEST, you find the following entry:

Pyroelectricity 2003
UF Pyroelectric coefficient
BT Electrical properties
Thermodynamic properties
Thermophysical properties
RT -Piezoelectricity

The number 2003 indicates that Pyroelectricity as all TEST terms belongs to a category group, in this case 2003 Electricity and magnetism. This group contains about 200 terms and is one of the 14 subdivisions of category field 20, Physics. There are altogether 22 fields divided into 177 groups. We further see that the term should be used for (UF) Pyroelectric coefficient, that the thesaurus contains three broader terms (BT) but no narrower term (NT). Finally, there is also a related term (RT), the hyphen indicating that this term is subdivided and may be worth while looking up.

At the end of TEST, after the Subject category index, we find a Hierarchical index, containing some 700 short hierarchies. These hierarchies are headed by terms which have no broader terms and at least two levels of narrower terms. They are built entirely on the indications BT and NT in the main section of TEST, and by no means all terms are included in a hierarchy.

The term Pyroelectricity is found in two hierarchies, Electrical properties and Thermodynamic properties. In the first of these it appears as a direct subdivision of the highest term, in the second we have the hierarchical order Thermodynamic properties - Thermophysical properties - Pyroelectricity.

The term Thermodynamics is not indicated as a BT to Thermodynamic properties, and does in fact not occur in any hierarchy. On the other hand, there is a category group 2013 Thermodynamics, which contains Thermodynamic properties and Thermophysical properties among its more than 100 terms, but not Pyroelectricity. We have here an instance of the fundamental difference between UDC and TEST hierarchies. There are many cases in the UDC where one and the same term appears in more than one of the UDC hieararchies. But then there is always the question of a difference in meaning or at least a different aspect. Thus we find Iron as a chemical element in 546 and as a technical material in 669. Pyroelectricity, on the other hand, is a definite physical property, and its meaning is exactly the same as a subdivision to Thermodynamic properties as to Electrical properties. In UDC it can appear in one place only, and, quite logically in our opinion, this place is in 537, Electricity.

The next example deals with the term Separation and some of its NT's. This term belongs to category group 1407, General concepts. Under this rather vague heading are found terms with widely different meanings. Among the narrower terms to Separation we find Precipitation (chemistry). This term belongs to the category group 0704 Physical and general chemistry. Several of the narrower terms belong to more than one category group, e.g. Electrostatic separation (1302 Civil Engineering and 1308 Industrial processes) and Distillation (0701 Chemical engineering and 1308 Industrial processes). Deasphalting and Flotation both belong to the same two groups, 0701 and 1308, the latter also to a third group, 1106 Metals. Further we find Isotope separation, which belongs to an entirely different group, 1802

Isotopes. In view of this diversity there is no wonder that the TEST hierarchy Separation doesn't bear any similarity to the UDC schedules, and the ASLIB group remark that the UDC codes often are scattered in the TEST hierarchies studied by them, should perhaps not always be interpreted as a criticism of UDC.

Sometimes the assignment of category groups to the TEST terms is not very consistent as shown by the following examples from the astronomical field. The category field 03 Astronomy and astrophysicsis divided into three groups, 0301 Astronomy, 0302 Astrophysics, and 0303 Celestial mechanics. It appears quite inexplicable that Dwarf stars belongs to 0301 and Giant stars to 0302, and that the terms Meteoroids and Nebulae belong to 0301 while their respective NT's Micrometeoroids and Planetary Nebulae have been referred to 0302. Further, while all planets belong to group 0301, Neptune alone has also been placed in group 0302. The astronomical terms are divided between several hierarchies, Astronomical instruments, Celestrial bodies, Orbits, Solar radiation, etc. Terms like Astronomical observatories, Radio sources (astronomy) and Quasars, however,have not found their way into any hierarchy.

Our last example deals with some metallurgical terms. The following is one of the shortest hierarchies in TEST:

- Steel making
  - Openhearth process
    - Acid openhearth process
    - Basic openhearth process

It does not require any special knowledge of metallurgy to understand that there are many other steel making processes than those carried out in an openhearth furnace, and that this hierarchy therefore is extremely unsatisfactory. It is therefore rather important to find out how other steel making processes are dealt with in TEST. We find the following instructions:

Bessemer process USE Bessemer converters
Electric furnace process USE Electric arc furnaces and Steel making
Basic Bessemer process USE Basic converters and Bessemer converters
Electric arc melting USE Arc melting

There are terms for many types of melting, all pertaining to steel, e.g. Induction melting, Openhearth melting, Oxygen melting, Vacuum melting, and also many terms for furnaces, e.g. Acid converters, Electric arc furnaces, Kaldo covverters, Vacuum furnaces.

With regard to their treatment in TEST the steelmaking methods could evidently be divided into three groups:

1. All converter processes are expressed by precoordinated terms for Converters. Sometimes two such terms are coordinated.
2. The terms for openhearth methods contain the word Process. The terms openhearth melting and openhearth furnaces are also allowed.
3. Other steel making processes are indexed by coordinating the term for the specific furnace employed with the term Steel making. In this group also, we have some terms containing the word Melting.

   Why the terms in the first group are not coordinated with the term Steel making as in group 3, is difficult to understand.

The terms for various types of metallurgical furnaces are divided between three different category groups, namely

1106	Metals
1301	Airconditioning, heating, lighting and ventilating
1309	Machinery, tools and industrial equipment

The various types of converters are divided between 1106 and 1309 in a completely inexplicable way. In the former category group we find Basic, Bottom blown and Oxygen blown converters, in the second Acid, Top blown and Air blown. Similar inconsistencies are found among other types of metallurgical furnaces.

To sum up, we have found both the category groups and the hierarchies rather disappointing from the point of view of a comparison between TEST and UDC, especially if one wants to use such a comparison for improving the UDC structure. A good many terms are scattered between category groups in a completely haphasard way. Some hierarchial groups are absurdly incomplete, others are headed by a term of such a general meaning that the subsidiary terms or groups of terms must logically be scattered between widely spread parts of any general hierarchical system.

3.2. Classification structures in MeSH and UDC.

MeSH has been constructed to contain terms of importance to the biomedical literature. The descriptors are used both for the indexing of the journal articles included in MEDLARS and for the assignment of subject headings to the monographs, reports, proceedings etc. contained in the monthly Current Catalog, also produced by National Library of Medicine, USA. The scope of MeSH is in fact much broader than could perhaps be assumed from the statement that it deals primarily with biomedicine. Even a brief look at the names of the categories into which the descriptors have been allocated, shows that the terms included refer to familiar biomedical topics as well as to biology in a broader sense and also to fields such as Psychology, Sociology, Technology, Humanities etc. This means that there can be no full concordance with UDC class 61. An information retrieval system devoted to biomedicine evidently must have a vocabulary which contains terms from many marginal fields, a finding pertinent to concordance problems with respect to UDC and any "specialized" thesaurus. The full categorization schedule in MeSH is displayed in table 10.

The subdivisions in UDC 61 are discipline based. MeSH does not subdivide systematically along the same lines, and therefore the study of basic principles of subdivisions in these two systems does not lead very far. MeSH should not be considered as a classification system in any strict sense of the word. It could, however, probably be put in the group of classified thesauri, mentioned in the introduction to the above mentioned ASLIB study. The whole classificatory structure in MeSH, as displayed in the Tree Structures, is only to be considered as a convenient tool for facilitating literature retrieval. Therefore principles of subdivision are of a pragmatic character, closely tied to the functional characteristics of MEDLARS. In this study we concentrated our attention on the practical task of looking more

Table 10.

MEDICAL SUBJECT HEADINGS
Categories and Subcategories

A	**Anatomical Terms**	218
A1	Parts of the Body	218
A2	Musculoskeletal System	220
A3	Digestive System	222
A4	Respiratory System	224
A5	Urogenital System	226
A6	Endocrine System	228
A7	Cardiovascular System	230
A8	Nervous System	232
A9	Sense Organs	236
A10	Tissues, Embryonic Structures	238
A11	Cells and Cell Constituents	240
A12	Body Fluids and Related Substances	242
A13	Terminology of Animals	244
B	Organisms	246
B1	Animal Kingdom - Invertebrates	246
B2	Animal Kingdom - Vertebrates	248
B3	Plant Kingdom - Bacteria	250
B4	Plant Kingdom - Viruses	252
B5	Plant Kingdom - Fungi, Molds and other Lower Plants	254
B6	Plant Kingdom - Mosses, Ferns and Higher Plants	256
C	Diseases	258
C1	Infectious Diseases	258
C2	Neoplasms, Cysts and Polyps	264
C3	Musculoskeletal Diseases	270
C4	Digestive System Diseases	274
C5	Respiratory Tract Diseases	278
C6	Urogenital System Diseases	280
C7	Endocrine Diseases	284
C8	Cardiovascular Diseases	286
C9	Hemic and Lymphatic Diseases	290
C10	Nervous System Diseases	292
C11	Sense Organ Diseases	296
C12	Skin Diseases	298
C13	Diseases of Nutrition and Metabolism	300
C14	Injury, Poisoning and Immunologic Disease	302
C15	Diseases Exclusively of Animals	306
C16	Neonatal Diseases and Abnormalities	308
C17	Symptoms and General Pathology	310
D	Chemicals and Drugs	312
D1	Inorganic Chemicals - Elements, Simple Compounds, and Related Terms	312
D2	Organic Chemicals and Structural Groups	316
D3	Anti-Infective Agents and Pesticides	326
D4	Antineoplastic Agents and Immunosuppressive Agents	330
D5	Autonomic Drugs, Cardiovascular Agents, Muscle Relaxants	332
D6	CNS Drugs, Antiemetics, Antihistaminics, Antitussive Agents	336
D7	Hematologic Agents, Gastrointestinal Agents, Agents for Fluid Therapy	340
D8	Hormones, Precursors, Metabolites, Substitutes and Antagonists	342
D9	Enzymes, Coenzymes, Enzyme Inhibitors and Precursors	344
D10	Amino Acids, Peptides, Proteins and Nucleic Acids	348
D11	Carbohydrates, Lipids, Vitamins and Related Compounds	352
D12	Immunologic Factors, Biological Factors and Substances	356
D13	Miscellaneous Chemicals and Drugs	358
E	Analytical, Diagnostic and Therapeut[illegible] Technics and Equipment	362
E1	Diagnostic Technics and Equipment	362
E2	Preventive and Therapeutic Technics and Equipment	366
E3	Anesthetic Tec[illegible] and Equipment	368
E4	Surgical Technics and Equipment	370
E5	Miscellaneous Technics, Methods and Equipment	374
E[illegible]	Dentistry	376
F	Psychiatry and Psychology	378
F1	Psychologic Mechanisms and Processes	378
F2	Behavioral Symptoms and Mental Disorders	384
F3	Behavioral Sciences, Psychological Tests, Psychotherapy, Services	388
G	Biological Sciences	390
G1	Biological Sciences and Biological Phenomena	390
G2	Health Occupations and Disciplines	396
G3	Environmental Health, Hygiene and Preve[illegible]edicine	398
H	Physical Sciences	400
I	Anthropology, Education, Sociology and Social Phenomena	404
J	Technology, Commerce and Industry	408
K	Humanities	410
L	Communication, Library Science and Documentation	412
M	Named Groups of Persons	414
N	Health Care	416
N1	Population Characteristics	416
N2	Health Facilities, Manpower and Services	418
N3	Economics, Organizations, Social Control	422
N4	Health Services Administration	424

carefully into only two MeSH categories, which were compared with corresponding structures in UDC 61. The categories chosen were those pertaining to anatomical terms and disease terms, both of which were judged to be of central importance to any biomedical classification system.

3.2.1. Anatomical terms

Anatomy represents a well organized body of knowledge where the basic principles of classification of terms according to organ systems can hardly be disputed. As could therefore have been predicted, MeSH and UDC show much similarity in their respective structures of the anatomical terms.

There are, however, some interesting divergencies even in this field. MeSH contains 13 subcategories within Category A, Anatomical terms. In UDC the subclass 611 is devoted to Anatomy. The two systems were first compared by a tabulation of the number of terms found in the corresponding subsections (table 11). If, in MeSH, the same descriptor was found in more than one hierarchial substructure within the same subcategory it was counted only once in that subcategory, whereas it was counted again if it appeared in another subcategory. (In this way the actual number of MeSH terms in Anatomy is somewhat overrated.) The total number of anatomical terms in the two systems differ. UDC 611.1/.9 has about 250 terms more than the corresponding sections of MeSH, A1 through A9. The number of terms in UDC corresponding to MeSH A10 through A13 were only partially countable for reasons which will be discussed later. Only minor differences were noted with respect to the classifications in these sections. Skin terms have been assigned different places in the two systems. In MeSH they are found in the subcategory for terms denoting parts of the body. In UDC they are placed in 611.7 together with terms for the muscul skeletal system. The hematopoietic and lymphatic systems terms are included in the subcategory A7 for the cardiovascular system terms in MeSH. In UDC they are grouped together with the terms for the ductless glands in 611.4.

Table 11. Anatomical subcategories in MeSH compared with UDC 611.

		Number of terms			
		MeSH	UDC		
A1.	Parts of the body (including skin: 8 terms)	50	30	611.9	Parts of the body
A2.	Musculoskeletal system	121	305	611.7	Musculoskeletal system (including skin: 20 terms)
A3.	Digestive system	77	37	611.3	Digestive system
A4.	Respiratory system	23	14	611.2	Respiratory system
A5.	Urogenital system	39	62	611.6	Genito-urinary system
A6.	Endocrine system	15	21	611.4	Lymphatic system. Haematopoietic system. Ductless glands. Chromaffin system. (endocrine syst.: 11 terms)
A7.	Cardiovascular system(including hematopoietic and lymphatic systems: 8 terms)	91	101	611.1	Vascular system
A8.	Nervous system	154	232	611.81/.83 + 611.89	Nervous system
A9.	Sense organs	37	48	611.84/.88	Sense organs
Sum of terms corresponding to MeSH A1 through A9		607	850		
A10.	Tissues, Embryonic structures	55	58	611-013.3/.8 611-018.2/.8	selected terms for embryologic structures selected terms for tissues
A11.	Cells and cell constituents	102	27	611-013.1 611-018.1 611-018.2/.8	Germ cells The cell selected terms for specific cells and subcellular structures
A12.	Body fluids and related substances	39	4+?	611-013 + 611-018	selected terms Main entries in 612 and 616
A13.	Terminology of animals	36	2+?	611.28/.29	Main entry in 59
Sum of terms corresponding to MeSH A10 through A13		232	91+?		

Table 12. Absolute numbers and relative distribution of anatomical terms in MeSH subcategories A1-A9 compared with UDC 611 terms.

	%		Abs. numbers	
	MeSH	UDC	MeSH	UDC
A1 Parts of the body	6,9	3,5	42	30
A2 Musculoskeletal system	19,9	33,5	121	285
X Skin	1,3	2,3	8	20
A3 Digestive system	12,7	4,3	77	37
A4 Respiratory system	3,8	1,6	23	14
A5 Urogenital system	6,4	7,3	39	62
A6 Endocrine system	2,4	1,3	15	11
A7 Cardiovascular system (including hematopoietic and lymphatic systems)	15,0	13,1	91	111
A8 Nervous system	25,3	27,3	154	232
A9 Sense organs	6,1	5,6	37	48
Sum: ≈	99,8%	99,8%	607	850

Table 12 shows the differences in the elaborations of details with regard to the various organs systems in absolute numbers as well as the relative percentages. For the sake of simplici in comparing, this list was slightly reorganized in order to separate out skin terms and also to allocate terms for the hematopoietic and lymphatic systems to the cardiovascular system terms. Some interesting differences show up here. MeSH has four quantitatively dominating groups of terms, (each with more than 10 % of the anatomical terms) in ranking order, the terms for 1/ the Nervous system, 2/ the Musculoskeletal system, 3/ the Cardiovascular system and 4/ the Digestive system. UDC has another ranking profile, the top three groups being, 1/ the Musculoskeletal system, (with a strikingly high percentage of anatomical terms) 2/ the Nervous system, and 3/ the Cardiovascular system. (Number four in UDC 611 is found to be the class for urogenital terms.) UDC differs strikingly from MeSH in the low number of terms, both absolutely and relatively in the class for the digestive system.

Table 13. Term usage in MEDLARS in comparison with number of terms in MeSH A1-A9 and UDC 611.1/.9 (Three year period: 1968-1970)

	Index Medicus, NLM		Terms	Terms
	Usage	% usage	MeSH	UDC %
A1. Parts of the body (Including skin)	40 092	7,9	8,2	5,8
A2. Musculoskeletal system	68 677	13,7	19,9	33,5
A3. Digestive system	81 083	16,0	12,7	4,3
A4. Respiratory system	19 717	3,9	3,8	1,6
A5. Urogenital system	57 250	11,3	6,4	7,3
A6. Endocrine system	31 548	6,2	2,4	1,3
A7. Cardiovascular system	92 279	18,2	15,0	13,1
A8. Nervous system	89 372	17,8	25,3	27,3
A9. Sense organs	24 472	4,8	6,1	5,6
Sum:	504 445	99,8%	99,8%	99,8%

In this connection it is of interest to look at the figures from the MEDLARS usage of MeSH terms. From a statistical analysis, performed at NLM, it can be calculated that the terms within the subcategories A1-A9 were used more than half a million times during the three year period 1968-1970. Percentages of this usage with respect to the subcategories A1-A9 are tabulated in table 13 and can be compared with the percentage distribution of the number of terms within the same subcategories. The interpretation of the relations between term usage and the number of terms has many pitfalls. One very important factor here is the fact that MEDLARS because of its indexing rules makes use of the anatomical terms in contexts outside the purely anatomical field, e.g. for indexing Physiology, Pathology and Pharmacology. UDC would place these topics under numbers from other classes than 611. The tabulation of the MEDLARS usage of terms therefore has limited concordance value in the anatomical field.

The four subcategories in MeSH with the highest number of terms also show the heaviest usage. The usage ranking order is not quite the same however, being, by names of subcategories: 1/ the Cardiovascular system, 2/ the Nervous system, 3/ the Digestive system and 4/ the Musculoskeletal system. The MEDLARS usage of terms within the subcategories for the Urogenital and Endocrine systems show higher actual use than the mere number of terms would betray. Looking at UDC in this connection the high number of terms for the Musculoskeletal system in 611.7 is still more pronounced in relation to the MEDLARS' use of these terms. For reasons mentioned above, there can be no true comparison between UDC and term usage in MEDLARS unless at least class 612 is also included. The rather extensive literature in Gastroenterology and Endocrinology for instance, would have to be classed primarily in 612. A marginal observation is that UDC seems more in balance with MEDLARS with respect to terms for the Urogenital system.

More detailed examinations with regard to the relations between the number of terms and the actual usage of terms would probably reveal many interesting facts. As mentioned earlier, the concordance problems between MeSH and UDC can not rely only on term statistics. The syntactic differences do in fact constitute the most important and perhaps the most challenging part of the study. Some syntactic problems become visible already in the discrepancies noted between UDC 611 and some MeSH subcategories in Anatomy, A10-A13, which must now be discussed. The last two of these subcategories A12 and A13 cannot really be compared with UDC 611 schedules. Two odd animal terms were noted in 611, Gills and Swimming bladder. Animal anatomical terms definitely belong to other parts of UDC and are therefore not further discussed here. The problems of body fluid terms are more complex. In 611 there are four such terms scattered among the auxiliaries, i.e. Amniotic fluid, Blood, Blood plasma and Embryonic blood. In 612 and in 616 more terms corresponding to A12 terms can be found. On examination, these terms proved to be rather confusing and interwoven with complicated syntactic structures and they will not be further pursued here. The subcategory A10 in MeSH provides a place for terms for tissues and embryonic structures. These terms are partly matched by the terms found in the special auxiliary subdivisions -013 and -018 in UDC 611. UDC is somewhat richer in terms for embryologic structures and tissues than MeSH. In A11 MeSH in better provided with terms for specific cells and subcellullar structures, whereas UDC betrays a particularly poor classificatory structure here. The syntactical problems touched upon in the examination of these last two subcategories will be elaborated in another context later in this paper.

3.2.2. Disease terms

The disease terminology in MeSH is found in the 17 subcategories in category C, Diseases. In UDC subclass 616, Pathology, contains the majority of the disease terms. A few terms are also found 613 and 614 and a considerable number in 617 and 618. Characteristic for the disease classification in UDC is,that there are very few main numbers for specific diseases, except in one of the sections for infectious diseases, 616.9. Two very elaborate auxiliary schedules, -001/-009 and to some extent also -01/-09, do in fact constitute the basis of the whole system. Therefore practically all main numbers in 616.1/.8 look like elaborate anatomical schedules, where the notation, however, has to be read "disease of organ so or so". More than 50 percent of the space in 616 is taken up by various examples of how to use the auxiliaries in order to construct adequate notations for specific diseases. A consequence of this is that it is impossible to count the number of disease terms in UDC,as the examples given are not meant to be exhaustive. In the sampling of terms from all subcategories in MeSH for the concordance study already discussed, we collected 34 terms from the 17 subcategories containing disease terms. It was found that it was possible in only 6 cases to assign only a main number from UDC. For the remaining terms the use of auxiliaries came into play with varying degrees of complexity. The very special devices in UDC with regard to disease classification warrants an excursion to make clear the difficulties facing those not familiar with the system. If a concept corresponding to the MeSH disease term could not be found in the alfabetical index and there was no example given in the main schedules, we had to construct the notations by using our knowledge of the characteristics of this disease or by consulting medical textbooks or dictionnaries. The fundamental difficulty with the 616 schedules is that there can be no consistent classification unless one elaborates private tools for the control of the system. It also seems necessary to make extensive use of A/Z devices.

These facts can be easily illustrated by discussing some problems met with in our small sample. For Sarcoidosis, which is a well defined clinical entity, also known as Mb. Besnier-Boeck-Shaumann, we could not find any explicit classification number. Sarcoidosis is a chronic disease, characterized by typical granulomatous changes in many organs, but specifically engaging the cells of the reticuloendothelial system. Either one could use the very simple notation 616-002.7, which stands for "Diseases characterized by granulomatous and nodular inflammation in general" or one could add Sarcoidosis as an A/Z device. Alternatively one could try to be more specific and use the notation 616.42-002.7, whereby the notation should read "Disease of the reticuloendothelial system characterized by ..." Even here one could of course consider adding the word Sarcoidosis for the sake of specificity.

A more serious case perhaps, is illustrated by the difficulties in classifying Froelich's syndrome. Here it is necessary to use the A/Z device because the syndrome cannot be uniquely defined otherwise. In classifying the same syndrome under the synonymous name Dystrofia adiposogenitalis one could construct the following notation 616.43 -07: [616-056.52 + 616-039.12],which should read "Endocrine disease with symptoms of obesity and sexual immaturity" Here it is unavoidable to add the name Froelich because it could not otherwise be separated from some other endocrine syndromes e.g. Cushing's syndrome[+], which among other symptoms also can present these two. In fact one could also classify Froelich's syndrome with regard to the fact that it can be a neurologically

---

[+] **Cushing's syndrome is sometimes given the notation 616.432-008.61 however. This is an easier notation but the meaning of the notation is not scientifically correct, being "hyperfunction of the pituitary gland". (Personal communication)** important problem, as symptoms such as headache, epileptic seizures and diabetes insipidus can present themselves in Froelich cases. One could also more specifically classify it as related to a tumor in the hypophyseal region etc. There are no rules or guidelines given in UDC for the many cases when the classification of a disease entity is not obvious. The whole apparatus of auxiliaries invites to the construction of notations which in many cases can be roughly descriptive of disease characteristics. But the danger obviously is,that these descriptions are either not precise or exhaustive or that they are scientifically unsound. This is especially obvious for diseases of unknown etiology and/or complex symptomatology. A thorough analysis of the consequences of the principles for disease classification in 616 seems to be a very necessary task for the revision of the 61 schedules.

In spite of the obvious difficulties involved in the study of the 616 schedules,we wanted to obtain an approximate figure for the number of disease terms in 616. A rough count was made, and all main numbers as well as all examples of the usage of -001/009 were included. Table 14.

Table 14. Estimation of the number of notations for diseases in 616 by main numbers and -001/-009 auxiliaries.

	Main numbers	-001/-009
616.1	400	100
.2	190	240
.3	110	380
.4	30	60
.5	90	180
.6	130	210
.7	430	260
.8	300	520
.9	140	-
	1820	1950

As the -01/09 auxiliaries were not included, the number of disease entities in 616 exceeds the approximately 3770 disease concepts found by the count conducted as described above. It was specifically difficult to include the -01/-09 series of auxiliaries. The notations in this series are mostly used as syntactic relators. The use is not logically consistent, however, as many notations within this series have another role, namely to aid in specifying a disease entity. By extrapolation the total number of disease entities and pathological concepts given or exemplified in UDC was estimated to exceed 4000. This means that UDC has at least 1200 more disease terms than MeSH. As no satisfactory concordance could be constructed to elucidate the number of terms for the various disease categories in UDC, table 15 is not fully parallel to the Anatomy concordance table. Only the structuring of disease terms into subcategories in MeSH is compared with the corresponding sections in UDC.

In table 16 the statistics of the MeSH disease terminology is shown. (The terms were counted according to the same criteria as the anatomical terms. Disease terms are allocated more often than anatomical terms to more than one subcategory, however. Therefore the total number of terms by categories in this count exceeds the sum of individual terms more sharply than in the Anatomy count). In table 16 the corresponding MEDLARS' usage figures are also shown. The number of terms for infectious diseases lead the MeSH statistics as to absolute number of terms, closely followed by terms for neoplasms. The subcategories next in ranking order with respect to number of terms are 3/ Digestive system diseases, 4/ Nervous system diseases and 5/ Urogenital system diseases. As was the case with the anatomy terms, the actual usage pattern differs from the term number profile. Here the ranking order for disease subcategories is 1/ Neoplasms, 2/ Cardiovascular diseases, 3/ Infectious diseases, 4/ Digestive system diseases. The ranking order between 1/ and 2/ might be questionable because indexing rules call for two neoplasm terms, one for the histologic type and one for the organ involved to account for one neoplasm entity. (It can be noted that here there is an interesting concordance between MeSH and UDC because UDC applies an auxiliary for histologic type to a main number for organ involved, thereby coding exactly the same information as MEDLARS about neoplasms).

Table 15. Subcategories in MeSH C1-C17 and corresponding sections in UDC.

C1.	Infectious diseases	616.9	Infectious diseases
		616-002.5/.9	"Specific infectious diseases"
		616-022.7	Diseases caused by bacteria, microorganisms ...
C2.	Neoplasms, cysts and polypes	616-006	Neoplasms
C3.	Musculoskeletal diseases	616.7	Pathology of locomotion
C4.	Digestive system diseases	616.3	Digestive system diseases
C5.	Respiratory tract diseases	616.2	Respiratory system diseases
C6.	Urogenital system diseases	616.6	Genito-urinary system diseases
C7.	Endocrine diseases	616.43	Endocrine diseases
C8.	Cardiovascular diseases	616.1	Pathology of the circulation and blood vessels.(excluding 616.15)
C9.	Hemic and lymphatic diseases	616.41/.42	Diseases of the haematopoietic and lymphatic systems.
		616.15	Diseases of the blood
C10.	Nervous system diseases	616.8	Neurology
C11.	Sense organ diseases	616.214.8	Olfactory diseases
		616.28	Otology
		617.7	Ophthalmology
C12.	Skin diseases	616.5	Dermatology
C13.	Diseases of nutrition and metabolism	616.39	Dietetic diseases
		616-008.9	Disturbances of metabolism
C14.	Injury, Poisoning and Immunologic disease	616-001	Traumata
		616-099	Poisoning
		616-097	Antigens, antibodies
C15.	Diseases exclusively of animals	619	Veterinary Medicine
C16.	Neonatal diseases and abnormalities	616-053.31/.35	Neonatal diseases
		616-007	Anomalies
C17.	Symptoms and General Pathology		Scattered auxiliaries

Table 15 is not complete as too many details would blur the overall picture. Some comments to this table are given in the discussion on problems of disease classifications.

Table 16. Term usage in MEDLARS in comparison with number of terms in MeSH C1-C17. Three year period: 1968-1970.

		MEDLARS usage of terms		Number of MeSH terms	
		abs. numbers	%	abs. numbers	%
C1.	Infectious diseases	123 358	9.9	382	13.9
C2.	Neoplasms	205 772	16.4	319	11.5
C3.	Musculoskeletal diseases	63 380	5.1	131	4.7
C4.	Digestive system diseases	109 122	8.7	224	8.1
C5.	Respiratory tract diseases	74 126	5.9	103	3.8
C6.	Urogenital system diseases	82 848	6.6	178	6.4
C7.	Endocrine diseases	37 133	3.0	70	2.6
C8.	Cardiovascular diseases	125 354	10.1	127	4.5
C9.	Hemic and lymphatic diseases	46 503	3.7	108	3.9
C10.	Nervous system diseases	95 377	7.6	204	7.3
C11.	Sense organ diseases	39 579	3.2	101	3.6
C12.	Skin diseases	35 074	2.7	144	5.2
C13.	Diseases of nutrition and metabolism	59 491	4.8	128	4.7
C14.	Injury, Poisoning and Immunologic diseases	84 348	6.7	160	5.7
C15.	Diseases exclusively of animals	19 182	1.5	88	3.1
C16.	Neonatal diseases and Abnormalities	40 627	3.3	127	4.5
C17.	Symptoms and general pathology	11 491	0.9	169	6.1
Sum:		1 252 765	100.1	2 763	99.6

The discrepancy between the elaboration of terms and the relatively low use in C17 can also be explained by referring to indexing rules. The low use of the terms for animal diseases is explained by the fact that MEDLARS indexes Veterinary Medicine only marginally and not with the ambition to cover this field comprehensively.

As it turned out to be very difficult to compare MeSH and UDC disease terms quantitatively, only some special observations will be added to what has already been said.

The classification of infectious diseases has about 140 main numbers in 616.9. There are also some special auxiliary schedules in 616-002.5/.9 for tuberculosis, syphilis and some other infectious diseases. 616-022.7 is another series for infectious diseases, where subdivisions of 576.85 are necessary to specify the infectious agents. This means that the classification of infectious diseases makes use of at least three different notational devices which, at least notationally, does not seem to be a good solution to this problem. Terms for infectious diseases are all held together in one subcategory in MeSH, which makes it much more easy to control the terms. The fact that MeSH has a rather well developed vocabulary in this field and the finding that the MEDLARS' use of these terms is quite impressive, makes it reasonable to conclude that in handling the classification of infectious diseases UDC could consider some changes. Another observation is that the development of the neoplasm vocabulary seems rather poor in UDC. The auxiliary schedules 616-006 contain approximately 120 neoplasm notations. The heavy use of neoplasm terms in MEDLARS and the importance of this field, definitely indicate the need for improvements in UDC.

It should be noted that the above mentioned observations pertain to two types of diseases, which are classified according to special principles, by etiologic agent and histopathologic type respectively. The difficulties with the classification principles in the main schedules in 616 have already been discussed. It can be added that MeSH contains the same sort of non-specific terms as the main numbers in 616.1/.8. In contrast to UDC,these non-specific disease terms constitute only a minority of the disease terms in MeSH, and special indexing rules govern the use of these terms.

Some disease subcategories in MeSH have no clear corresponding sections in UDC. Generalized diseases for instance, like metabolic diseases and immunologic diseases can only be accounted for by use of auxiliaries. Lupus erythematosus, LE, is an example of a generalized disease. It was looked up in the index to 616. Here one finds one notation 616.5-002.73 for Lupus erythematodes. The notation should read "Skin diseases with leprous inflammation". This cannot be the looked for notation as LE is not skin lepra. The index also says Lupus erythematosus. For this entry two notations are given. One is 616-056.3:616.5-002.524, which should read "Allergic disease of the skin in tuberculosis". The other notation is 616.5-002.525.2,which is another notation saying "Skin disease in paratuberculosis". As LE is not a tuberculous disease no adequate notation is therefore found. This example may show 616 at its worst. It was accidentally found by looking for an example of a generalized disease, but as it also illustrates what has been said before,this finding leads to serious misgivings as to the consequences of the 616 classificatory principles. LE is called 'Lupus Erythematosus, Systemic' in MeSH and is found in a group of diseases named collagen diseases in C17. For this broader concept there is no correspondance in UDC.The classification of LE in UDC reflects the fact that historically speaking LE was not separated out as disease entity of its own until rather lately. It was considered to be a special kind of tuberculosis or may be even manifestation of lepra. The name lupus, meaning wolf, is an old name for disfiguring skin diseases, very often demonstrated to be of tuberculous origin. As LE was separated out as a disease entity, part of the old name was retained, as has often been the case in the development of medical terminology. Lupus erythematodes and Lupus erythematosus today are synonyms, however, for this new disease entity. UDC also in this respect adds to the confusion by treating them as separate diseases.

The general impression is that UDC 616 in many respects is antiquated. Even a superficial examination of the 616 schedules reveals serious shortcomings. It is not fair, however, to compare UDC with MeSH for many reasons. MeSH has the advantage of being rather recently compiled and of being subject to continous vocabulary control by specialists. The backbone of the 616 schedules, and the classificatory principles with heavy reliance on auxiliaries have resulted in a structure which inevitably,as it seems, has led to an unhappy conservation of old ideas, illustrated in the case of Lupus Erythematosus. Other examples could probably also be found. The revision of 616, which has to be done in order to bring this subclass up-to-date, therefore,in our opinion, has to start with a discussion of more suitable principles for classification of diseases.

4. Syntactic structures

In the preceding paragraphs syntactic problems were mentioned in various connections. Here we will try to deal more specifically with some of these problems. It is evident that the UDC 61 schedules contain many notations capable of expressing relations between terms and also notations for aspect specifications of terms. This is what we refer to as the syntactic structures. MeSH also has a special type of operators for these purposes, the 60 subheadings already mentioned. As the auxiliary tables in UDC in many instances are exactly equivalent to MeSH subheadings, a concordance to the MeSH subheadings was compiled. This is not presented here, since the real problems involved in comparing the two systems must lead to considerations of the totality of the paradigmatic structures with regard to the rules governing the applicability of the relators. In other words, a compilation of a subheading concordance is only a first step towards a syntax concordance and as such a necessary tool for the next step, which is to study the grammatical rules governing the use of the syntactical notations. In MeSH the rules for the applicability of subheadings constitute a very substantial part of the indexing manual, a fact which points to the crucial importance of these rules. In UDC there is nothing comparable to help the user to clarify these problems. Another important difference between the systems from the syntax point of view, is that the MeSH grammar was constructed to cover all terms in all categories, whereas there is very little evidence in UDC that such an overall view was ever attempted. The MeSH subheading grammar is built upon a consideration of applicability range of subheadings with regard to categories and sometimes subcategories. In table 17 the diagram plotting of subheadings against categories clearly shows the grammatical structure of the total system. An attempt was made to plot in the same way the auxiliaries of the series -01/-09 against the subclasses 611, 612, 615 and 616 in UDC, table 18. Within these four UDC subclasses the -01/-09 auxiliaries were estimated to give rise to over 1000 notations, table 19.

Table 17. MeSH subheading-category applicabilities

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Abnormalities	+													
Administration & dosage				+										
Adverse effects				+	+			+		+				
Analysis	+	+		+			+			+				
Anatomy & histology	+	+												
Antagonists & inhibitors				+										
Biosynthesis				+										
Blood			+	+		+								
Blood supply	+													
Cerebrospinal fluid			+	+		+								
Chemical synthesis				+										
Chemically induced			+			+								
Classification		+	+	+	+	+	+	+	+					+
Complications			+			+								
Congenital			+											
Cytology	+	+												
Diagnosis			+			+								
Diagnostic use				+				+						
Drug effects	+	+				+	+							
Drug therapy			+			+								
Education						+	+	+	+					+
Embryology	+	+	+											
Enzymology	+	+	+			+								
Etiology			+			+								
Familial & genetic			+			+								
Growth & development	+	+												
History			+	+	+	+	+	+	+	+	+	+	+	+
Immunology	+	+	+			+								
Injuries	+		+											
Innervation	+													
Instrumentation					+	+	+	+		+				
Isolation & purification		+		+										
Manpower					+	+	+	+	+			+		+
Metabolism	+	+	+	+		+								
Microbiology	+		+											
Mortality			+		+	+								
Nursing			+		+	+								
Occurrence			+			+								
Pathogenicity		+												
Pathology	+		+			+								
Pharmacodynamics				+										
Physiology	+	+		+										
Physiopathology	+		+			+								
Poisoning				+						+				
Prevention & control			+			+	+							
Radiation effects	+	+		+			+							
Radiography	+		+			+								
Radiotherapy			+			+								
Rehabilitation			+			+								
Secretion	+			+										
Standards				+	+	+		+	+	+		+		+
Supply & distribution				+	+							+		+
Surgery	+		+			+								
Therapeutic use				+				+						
Therapy			+			+								
Toxicity				+						+				
Transplantation	+													
Urine			+	+		+								
Utilization					+							+		+
Veterinary			+		+									

Table 18. Number of auxiliaries in the -01/-09 series with "equal meaning" applicability in more than one 61 subclass among subclasses 611, 612, 615 and 616.

	611	612[x)]	615	616	
-01	+	+			1
-018	+			+	1
-019	+	+			1
-031			+	+	1
-032			+	+	1
-033			+	+	1
-034			+	+	1
-035			+	+	1
-035.1/.9			+	+	6
-036			+	+	1
-036.2/.8			+	+	3
-06			+	+	1
-071			+	+	1
-073			+	+	1
-074			+	+	1
-08	+	+			1
-089		+		+	1
-099			+	+	1
	4	4	20	22	25

x) Both 612-0... and .0...

In this compilation no distinction has been made between auxiliaries .01/.09 and -01/-09 since the revisions of some of the 61 schedules have resulted in notational changes here without implying any specifically syntactic changes.

It is very clear that only a small fraction of these notations show the looked-for characteristics, namely of being employable in a way comparable to the majority of the subheadings in MeSH, i.e. carrying the same meaning in two or more subclasses.* The finding of only 25 such cases out of the over 1000 notations shows

---

* The syntactic functions of those subheadings which have either only one single category applicability or which can be applied to descriptors from more than five categories present special cases which will be analyzed and discussed in a second study devoted entirely to syntactic concordance problems.

Table 19. Estimation of number of -01/-09 auxiliaries in 611, 612, 615 and 616 (excluding special use in subsections in these subclasses).

	611	612 x)	615	616	Sum
-01	103	129	51	2	285
-02	-	-	-	139	139
-03	-	-	25	162	187
-04	-	-	-	-	-
-05	-	-	-	129	129
-06	15	3	3	1	22
-07	-	-	10	105	115
-08	1	7	-	166	174
-09	-	-	1	33	34
	119	139	90	737	1085

---

x) Both 612-0... and .0...

that the various subclasses in UDC 61 on the whole are not among themselves syntactically concordant. From table 19 it can be seen that the use of auxiliaries in the here studied four subclasses is uneven. Only in 616 is use made of the whole -01/-09 series. (-04 is empty).

Categories A, D and C in MeSH to some extent correspond to 611, 615 and 616 in UDC. A-category terms in MeSH can be syntactically specified by 22 different subheadings. Two of these can also be applied to D-category terms and eight to C-category terms. C-category terms in MeSH allow 29 different subheading combinations. Five of these subheadings are allowed also for D-category terms. D-category terms can have 23 subheadings, seven of which also go with A or C terms. The syntax concordance in 611, 615 and 616 in UDC has no similar pattern.

Only between 615 and 616 does there seem to be an interesting similarity. The syntactical functions involved are mostly very different from the MeSH functions, however. An enquiery into the special use of the .01/.09 auxiliaries in 612 reveals, that here the auxiliaries are more elaborated, and that for special purposes in 612 the full range of the .01/.09 series is used, Table 20.

Table 20. Special meaning of .01/.09 in subsections of 612

	.01	.02	.03	.04	.05	.06	.07	.08	.09	Sum
612.1										
.2		1						1		2
.3	4	1								5
.4	1	1	1	1	1		1	1	1	8
.5										
.6	1	2	5	1	5	1	1	1	1	18
.7										
.8	8		1		1			1		11
.9										
Sum	14	5	7	2	7	1	2	4	2	44

A disturbing finding, however, is that in 612, where the syntax seems to be rather elaborated, the main numbers in themselves very often signify syntactical complexes. This means that the auxiliaries in 612 have not been structured for systematic separation of entities and relators. As a practical consequence of these findings one must conclude that any revision of the UDC 61 schedules probably would benefit from a consideration of the syntactical structures. There can be no transparency in the schedules if the basic syntactic structures are not laid open. Problems of concordance with other systems can probably never be satisfactorily solved unless these structures are well defined and consistently worked out. At least with regard to MeSH and other systems with elaborate syntactical structures this seems to be a logical conclusion. Much more could be said about these problems in MeSH and UDC 61, but as a second study will be devoted to this theme we will not discuss it further here. As a final point we have instead chosen to illustrate the syntax concordance intricacies by a very simple example.

Apart from the usage of -013 and -018 as the basis for the construction of some embryology, histology and cytology hierarchies in UDC 611 it is also stated that-013 stands for "embryology of' and that -018 can be used to mean "histology of". It is not made clear in the same explicit way whether -018.1 can stand for "cytology of" but this can be assumed to be the case. In any case it seem clear that the notations for entities and aspects are not adequately separated in this way. By classifying for instance "cytology of cartilage" one could use the notation 611-018.3-018.1, where 611-018.3 stands for cartilage and -018.1 for "cytology of". Notationally this must be admitted to be very unsatisfactory.

MeSH subheadings for morphology aspects are three, /anatomy & histology, /cytology, and /pathology. UDC also has some morphology aspect auxiliaries in 616. The applicability of the notational devices for these aspects in the two systems can be seen in table 21. It is evident that MeSH has limited discriminatory powers in this area (Special restrictions as to the applicability even for these three subheadings are given in the indexing manual, but the intricacies with exceptions and special limitations will not be detailed here). As can be seen, MeSH does not allow for differentiation of morphology studies on different levels in disease studies by means of subheadings (here coordination with main headings is used).

Table 21. Syntactical devices for expressing morphological aspects of organs and tissues in health and in disease.

1. In health	MeSH	UDC
ANATOMY	/anatomy & histology	611.1/.9
HISTOLOGY	/anatomy & histology	611-018
CYTOLOGY	/cytology	611-018.1
2. In disease		
PATHOLOGY	/pathology	616-091
HISTOPATHOLOGY	/pathology	616-091.8 or 616-018
CYTOPATHOLOGY	/pathology	616-091.8 or 616-018.1

UDC has no special indicator for the macroanatomical level, which is assumed to be implicated in the main numbers from 611. A potential logical development of the -018 series would be to take care of still finer morphological levels that are covered by cytology, i.e. the molecular biology level, by a further subdivision of the -018 series. But here the UDC system is caught in the trap of not having made the logical differentiation between syntactical and paradigmatic notations mentioned above. It can also be seen from table 21 that UDC tries to differentiate more than MeSH with regard to disease morphology studies, but here one finds that there are two ways of treating histopathology and cytopathology, either according to the subdivisions of -018 in 611 or to use just the one notation from the auxiliaries in 616, whereby it is no longer possible to separate the two aspects. Both systems could become changed into compatibility and internal consistency by introducing some changes. They might result in a concordance table like the one displayed in table 22.

Table 22. Possible concordance construction of syntactical devices for morphological aspects of organs and tissues in health and in disease.

1. In health	MeSH	UDC
ANATOMY	/anatomy	611.1/.9 or 611-0X
HISTOLOGY	/histology	611-018 or 611-0X.1
CYTOLOGY	/cytology	611-018.1 or 611-0X.3
2. In disease		
PATHOLOGY	/pathology	616-091 or 616-0X
HISTOPATHOLOGY	/histology or /histopathology	616-018 or 616-0X.1
CYTOPATHOLOGY	/cytology or /cytopathology	616-018.1 or 616-0X.3

References

1. ENGINEERS JOINT COUNCIL. Thesaurus of Engineering and Scientific Terms. New York, 1967.
2. WELLISCH, H. Concordance between UDC and EJC/TEST. FID document C 71-10 (26 April 1971), presented to FID/CCC at their meeting 10-12 May 1971.
3. WELLISCH, H. and I. de VRIES. A Concordance between UDC and Thesaurus of Engineering and Scientific Terms (TEST). Results of a pilot project. Paper presented at the international symposium "UDC in Relation to other Indexing Languages", 28 June to 1 July, Herceg Novi, Yugoslavia.
4. NATIONAL LIBRARY of MEDICINE. Medical Subject Headings 1971. Washington D.C., 1971.
5. COMITE INTERNATIONAL POUR LA DOCUMENTATION DES SCIENCES SOCIALES. Thesaurus Documentaire d'Economie (Premier projet). 19 fasicules. Paris, 1969 - 1971.
6. FÖLDI, T. and K. SOMMER. Konkordanz zwischen Thesaurus documentaire d'économie und Universelle Dezimalklassification (fas. 8, 9, 14, and 17). Report to FID/C3 Arbeitsgruppe für Planung und Organisation 30 October 1970.
7. VICKERY, B. C., M. SLATER, A. PESANIS and R. REYNOLDS. Classification in Science Information. A comparative study undertaken by ASLIB for the International Council of Scientific Unions. London, June 1969. Presented as document UNISIST/CSI/5.8 at the Fourth Session of the Central Committee to study the feasibility of a world science information system at Unesco House, Paris, 15-17 December 1969.
8. ARNTZ, H. Die DK - eine Vielfacettenklassifikation. Nachr. Dok. 21 (1970):4, p. 139-142.
9. ENGINEERS JOINT COUNCIL. Thesaurus of Engineering Terms. New York, 1964.
10. FILL, K. Einführung in das Wesen der Dezimalklassifikation. 3. Aufl. Berlin 1969.

THE CONCEPT
"MAIN CLASS" AND ITS IMPLICATION FOR STRUCTURAL IMPROVEMENT OF UDC

Jean M. PERREAULT
Library University of Alabama in Huntsville, USA.

.0 SITUATION

.01 ORIGIN

UDC has always been an information language that has been a-tinkering. Otlet and La Fontaine started it off by tinkering with DC, and no other general classification has ever had so many cooks involved in the simmering and stirring of its broth.

.02 TWO TYPES OF TINKERERS

Most of those who sought either to modify its structure to allow it better to conform to the structure of the documents to be organized by means of it, or to render it more intelligible to users (classifiers and searchers alike), have done so from one of two points of view: Some are practitioners to whom theory is at least somewhat repugnant. Their attempts may be characterized by great courage, by a willingness to take on vast labors; but not uncommonly they may be too particular, insofar as they concentrate on a detail or a group of details under the bland assumption that the change they advocate will not upset the system as a whole. Others are more theoretically inclined, and among these I must surely number myself. Their flaw, at least in the eyes of the practitioner-tinkerers, may be so much that they see UDC through the eyes of outsiders as that they are theorists. And the eyeglasses through which most such theorists today (myself included) look at UDC is Ranganathan's theory of analytico-synthetic classification.

.03 PRACTICAL NEED

At my own institution, the Library of the University of Alabama in Huntsville, the shelf-classification is LC and the catalog-organization LCSH. Discussion there focusses on several cases that relate to the appropriateness and helpfulness (or lack thereof) of the locations of certain groups of material. a: Psychotherapy and some (not all) of psychopathology are located in R... (Medicine) while most of psychology and the rest of psychopathology are in BF... (Psychology). b: Meteorology, Terrestrial Magnetism, and certain other closely related matters are located in QC... (Physics), while most other "planetary" subjects are to be found making up notationally second-order groups such as QB... (Astronomy) and QE... (Geology). c: Anthropology, GN..., contains little but general matters relative to that science, with much of its application to ethnic or geographic foci classed in D.../F... (History and Description). To use a system of classification for browsing means to arrange materials with it; to be forced to indicate the division to which each document belongs by some symbol other than its class membership call the classification system into question in terms of its appropriateness for such use. The fact that these locatings make the establishment of a divisional plan (to include such groupings as Behavioral Sciences and Environmental Sciences) difficult, by removing from the divisional collection documents used in connection with others and placing them in a division made up of documents along with which they would not be used, may be only a practical disadvantage, but I see no significant distance between theory and practice in information retrieval. It is therefore a problem deserving a theoretical solution for the sake of practical improvement; or perhaps the solution should be in terms of a reexamination of the choice of LC by comparison to such other systems as UDC and CC. But what are their provisions for such groupings based on, and what are the service-advantages and -disadvantages of their structures?

.1 METHODOLOGY

.11 INDIGENOUS METHODOLOGY POSSIBLE?

Is it possible to work up a theory for UDC that does not rely upon such theories as Ranganathan's? I wish to try, at least, and the first two methodological points to raise (and to hope to observe in the performance) are (1) that theory is not something really distinct from practice, at least in this field, and (2) that description, as well as explanation, is theory, but that the former avoids the normative tendency of the latter.

.12 QUALIFICATIONS ON THESE METHODOLOGICAL POINTS

Taken unqualifiedly, these methodological points are too little focussed to be of much use.

.121 THE THEORETICAL AND THE PRACTICAL

The first must be qualified at least to the extent that use is accepted as fairly distinct from theory: the use, that is, which the classifier or the searcher makes of an information language. But theory in the sense of a grasp of the transparent structural principles of UDC or any other general classification is intimately bound up with the way this ideal structure is embodied in schedule, index, etc. Thus I do not see the classificationist as being less theoretic than the "pure" classification theorist; the latter is merely concerned with making nomological propositions (i.e., purely descriptive, but valid in a broadly general way, and embodied in categorical or conditional statements), whereas the former takes the same propositional content and embodies it in a mixture of implicitly reflexive and imperative statements. The situation is not vastly different from that in a natural language: The classificationist is the poet, he who speaks the language in its richest shape, who reflects on his own speaking in hope of raising its lucidity and power ever higher; the classifier is the poet's reader, who may never himself poetize, but who is shown by the speaking of the poet how to speak--or by the painter or sculptor how to see colors, shapes, etc.--; the searcher is the poet's reader's hearer. Where does the "pure" classification theorist stand in all this? He is the grammarian, who does not invent, as does the poet (in Greek, *poíēsis* is both *poetry* and *making*), but who generalizes the best practice in the language into nomological propositions.

.121+ NOTE

It must be admitted that the normal natural-language grammarian also indulges in a great deal of explicit imperatives--normative statements as well as nomological. The point of the whole discussion is, however, that theory and practice are not really distinct except at a level below that of the construction and analysis of actua classifications; tabularly:

PERSON	ACTIVITY	"LEVEL"
"pure" classification theorist	makes nomological statements	theoretical
classificationist	reflextively constructs normative statements	
classifier	constructs document-surrogates	practical
searcher	retrieves documents	

It is the shaded-in area of overlap between "theoretical" and "practical" that is at issue: it is that portion of the whole spectrum where there is no real distinction--even if there is one between what is above it and what is below it.

.122 TERMINOLOGICAL CLARIFICATION

The second methodological point must be qualified by a fuller clarificiation of terms: Theory, Description, Explanation, Nomological, Normative. In very large measure my usage is at least attempted as in conformity with that of Edmund Husserl's Logical Investigations. Theory is taken as generic to all the others, and is thus not identical with Explanation. Explanation is taken as having a definite standpoint: "the reason for this is that", which clearly implies judgments of value, causality, and/or priority outside the object being examined. A Nomological statement is of the form "it is the case that every A is B", or "wherever A, then B". A Normative statement is of the form "every occurrence of A shall be made to conform to B as well", or "wherever A, then B". The one that is left is the hardest: Description. Can there be Description without at least implicit Explanation?

.122+ NOTE: LAWS AND RULES

Note that Nomological statements cover, at least partially, both Description and Explanation, but that Normative statements are neither explanatory nor descriptive, but imperative; and also that, as asserted above, Theory covers Normative as well as Nomological statements. Tabularly:

KIND OF STATEMENT	KIND OF ACTIVITY	INTENT
nomological	description	grasp laws of behavior
normative	explanation	discover why laws apply & operate
	construction	enforce laws through norms (rules)

It must be noted with particular attention that "laws" are not the same as "rules"; laws are statements of fact (or: statements of value?), whereas rules are statements of the conditions for the embodiment of those values. It is laws that we hope to grasp through non-explanatory description.

.123 CONTINUATION: METHODOLOGICAL INADEQUACY

The question Can there be Description without at least implicit Explanation? could probably only be answered with an unqualified Yes if we could see an example of a successful attempt along those lines. Since I have indicated allegiance to one who has so attempted, I could legitimately be asked to affirm his success--but I cannot count the transcendental idealism of the Cartesian Meditations as a success from the point of view of the supposedly purely descriptive Logical Investigations.

Still, philosophically flawed as it may well be, I need some methodology, and have chosen this one.

.2 THE MATTER UNDER CONSIDERATION

.20 WHAT IS A MAIN CLASS?

The topic What is an m.c.? is one that receives even more attention in informal training of library users than it does in print. There is a tendency to consider it as too closely allied to What is the best collocation of m.c.s?to be taken any more seriously than the latter. But I think a great deal of structural and operational improvement might eventuate from a clearer picture of the m.c. as a structural and operational feature of a general classification.

.21 DISTINCTION BY NOTATION

.211 CASES IN UDC AND LC

The most superficial sort of answer to the question What is an m.c.? would be (in UDC) Every concept symbolized by 0..., 1..., 2..., or the like; or (in LC) Every concept symbolized by A..., B..., C..., or the like. Yet that is in fact a surprisingly common answer. What is really wrong with it? I put another question in return: If (in LC) P... symbolizes the m.c. Literature, how is it that the m.c. History is symbolized by C..., D..., E..., and F...? Or again, if (in UDC) 37... symbolizes the apparently non-m. c. Education, how is it that (in LC) Education is an m.c., symbolized by L...?

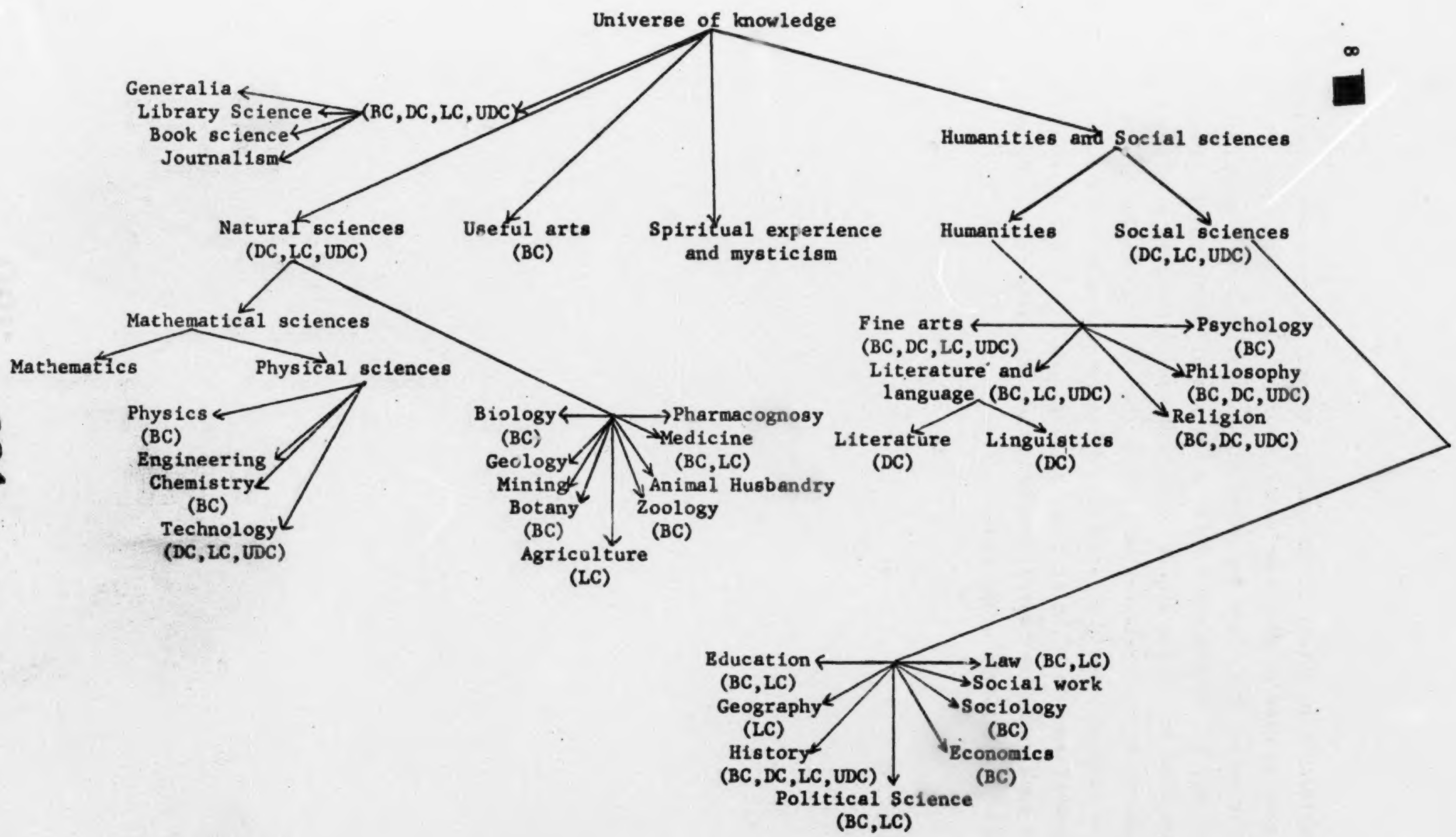
.212 DISCUSSION OF CASES IN UDC AND LC

In the Literature case, a reinforcing argument is that whereas there is a subject heading in the LC catalog: French literature, just as there is one: France--History, there is none: French philosophy. We have instead: Philosophy, French. Thus, whereas we see in a perfectly descriptive *thēoría* (view) that, in the subject headings, philosophy is taken as a unitary subject to be broken up by national characterizations, and, in the classification, is even pushed in with religion into "m.c." B..., we also see that in the subject headings literature and history are taken as so broad as to need immediate national qualifications. But the parallel breaks down there, presumably because history is even broader than literature, thus requiring a greater number of m.c.s so as to spread out use of digits evenly. This argument cannot stand up too well against the comparative size of the schedules concerned: 1630 pages for C..., D..., E..., and F... together, 2771 for P... alone.

In the Education case, comparison between two different systems, each with a different radix, can be taken as the reason: a longer radix allows more m.-c.-defining first-order symbols, so naturally where LC gives m.-c. status to education, UDC can not. Why, then, I ask in return, does UDC give m.-c. status to both Philosophy and Religion (1... and 2...), whereas LC confounds them into B...? An answer is easy enough to come by--but it is explanatory rather than descriptive and really gets us nowhere toward a true understanding--: because UDC derives from DC, which gave m.-c. status, etc., etc.

.213 COMPARISON OF SEVERAL CLASSIFICATIONS WITH CC

CC is a case (and BC might well also be; I do not intend to consider it in detail in this paper, though) in which the notational answer is made to make some sense. Not only is almost every first-order-symbol m.c. from any other classification here considered (BC, DC, LC, UDC) given a first-order symbol in CC whether or not more than one of the others do so, but Ranganathan has also provided summary- or super-classes at various (but not all) useful levels (the diagram shows the direction of subordination of m.c.s and super-classes in CC; next to each class name is the abbreviation for each other considered classification that also treats the same subject as a first-order-symbol m.c.):


Universe of knowledge
Generalia
Library Science
Book science
Journalism
(BC,DC,LC,UDC)
Natural sciences
(DC,LC,UDC)
Useful arts
(BC)
Spiritual experience
and mysticism
Humanities and Social sciences
Humanities
Social sciences
(DC,LC,UDC)
Mathematical sciences
Mathematics
Physical sciences
Physics
(BC)
Engineering
Chemistry
(BC)
Technology
(DC,LC,UDC)
Biology
(BC)
Geology
Mining
Botany
(BC)
Agriculture
(LC)
Pharmacognosy
Medicine
(BC,LC)
Animal Husbandry
Zoology
(BC)
Fine arts
(BC,DC,LC,UDC)
Literature and
language (BC,LC,UDC)
Literature
(DC)
Linguistics
(DC)
Psychology
(BC)
Philosophy
(BC,DC,UDC)
Religion
(BC,DC,UDC)
Education
(BC,LC)
Geography
(LC)
History
(BC,DC,LC,UDC)
Political Science
(BC,LC)
Law (BC,LC)
Social work
Sociology
(BC)
Economics
(BC)

(BC has one main class not found in any other: Astronomy and Earth sciences; LC has three such: Music, Military Science, and Naval Science.)

To the question What is an m.c.? the answer Every concept symbolized (in CC) by a..., 1..., . . . A..., AZ..., B..., or the like is an m.c.--is not then entirely superficial. The reason it does not make sense in the great majority of other cases is because in all the other classifications so many first-order-symbol "m.c.s" are congeries of m.c.s, i.e., super-classes of greater or lesser extension and usefulness.

.214 SECOND-ORDER CLASSES AND/OR SUBCLASSES

But if the question is asked How do you subdivide such a first-order-symbol "m.c." into its true constituent m.c.s?, the answer would be just as likely to be misleading in most cases if it were By finding the next level of symbolic display, i.e., the next notational order. Of course there are cases where such an approach can be fruitful; in fact it is almost invariably the case that the next level of notational division of a first-order-symbol "m.c." does in fact represent the primary divisions of that subject (or, if the first notational order represents a cluster of m.c.s, the m.c.s within it). What is unfortunately not true is that there is any way of distinguishing m.c.s on the basis of their notation alone, whether at the first- or second-order level of notation: in the second-order case this is so just because there is no ritual way to distinguish between a second-order m.c. and a second-order subclass, any more than between the m.c.s (in LC) North American History and Description (E31 - F1170) and Latin American History and Description (F1201 - F3799).

.3 DISTINCTION BY DISCIPLINE

.31 DISCIPLINES AND "CLASSIFICATION BY ATTRACTION"

A far less superficial approach, because it is based on meaning rather than on mere notational symbols, is to explain the formation of m.c.s by the attractive power of disciplinary specialism. It is of course regarded as one of the prime flaws in the work of a classifier if he "classifies by attraction", namely by what in the title or in his analysis of the document would most strongly attract the attention of the searcher, regardless of the orientation of the document as a whole. For instance, since the vast majority of documents on Jews and matters Jewish would fall (in DC) in class 296..., to "classify by attraction" is to put everything into 296... even if it deals with wholly non-religious aspects of matters Jewish.

.32 THINGS AND ASPECTS

Classification-construction by the attraction of disciplinary specialism is what is done in the case of all general classifications except Brown's Subject Classification; this is an attraction that purports to be wholly different from that which was called a flaw in §.31. That is to say, they group *things* together in terms of their *aspects*: the same things are dealt with in physics and in chemistry, in the main, or in religion and in cultural anthropology, in geography and in history, in zoology and in animal husbandry; the reason for the separation of the treatment of the nature of man in Religion

UDC	CC	LC
233...	Q:34...	BT700-745

and in cultural anthropology

UDC	CC	LC
39...	Y:1...	GN307-499

is that different academic specialists treat the same things as objects of their own specialisms. An important clue towards a final answer to our basic question is the fact that the animals treated both in zoology and animal husbandry are systematized by their relevant academic specialists in quite different patterns. A similar example is Mills' discussion[1] of the need for a different place-table under Law as against most other subjects.

.32+ NOTE: NON-IDENTITY OF EXTENSION

It must not, of course, be concealed that at least part of the extension of things dealt with in two m.c.s differentiable by different aspects of the same things may not be shared by both m.c.s: only

---

[1] Jack Mills, *A Modern Outline of Library Classification* (London, Chapman & Hall, 1960), p. 35.

DC 293

useful animals and those relevant to their usefulness (e.g., parasites) are dealt with in Animal Husbandry, whereas Zoology at least intends to deal with all animals whatever[2].

.33 HOW FAR "SPECIALISM" DOMINATES RECOGNITION OF DISCIPLINES

What level of generality/specificity then shall we set as the criterion for the recognition of an m.c.? (Recognition, here, not merely distinction, since the non-bibliographical world's recognition must be determined before we notationalize that recognition: it is the subject specialist's decision that matters here, not the librarian's Assuredly there are few theologians who are also cultural anthropologists; but even if some do exist, they are dual specialists. Nor does being a chess-master and an aerospace engineer simultaneously in any way assimilate these two specialisms. So we cannot apply the test of mutual exclusivity among the specialists themselves, any more than we can recognize m.c.s as distinct from the mutual exclusivity of their objects.

.34 CASES IN LC

Still, everyone would agree that History is an academically specialized discipline (even if its practitioners claim all things as their domain) wholly distinct _qua_ discipline from the disciplines of Physics or of Sociology. What about, within History taken in its very broadest sense, the question about m.c.s based on thing-distinction where the things concerned are places (Latin America, Africa, etc.), or belong respectively to metaphysically disparate categories (artifacts as the things dealt with in archaeology, recorded events as the things dealt with in history proper)? Again, everyone would agree that whereas it is the case that Zoology and Animal Husbandry are primarily distinct in terms of aspect (specialism) and secondarily so in terms of extension (objects dealt with), and that Religion and Cultural Anthropology share the same partial material and total formal distinction, it is also the case that material distinction can be (at least almost) total; witness that between the two mentioned pairs: Religion and Cultural Anthropology as a semi-unity share almost none of their objects with Zoology and Animal Husbandry on the other--excluding perhaps animals worshipped, totemized, and sacrificed--, Physics and Chemistry almost none with either Religion/Cultural Anthropology or Zoology/Animal Husbandry.

---

[2] _Cf._ Claude Levi-Strauss, _The Savage Mind_ (University of Chicago Press, 1966), pp. 138-139, etc., on the contrast between "scientific" and folk-classification.

.34+ NOTE: OBJECT AS COMPLEX OF THING AND ASPECT

There does arise a question which I shall avoid, both because this paper is not supposed to be philosophical but bibliothecological, and because I do not see myself as competent for its solution: How legitimately distinct are the two mentioned distictions, Formal and Material? Are the organic chemist and the chemical engineer not concerned with the same fats that the physiologist studies? Or is the aspect-distinction so strong as to suffice to make an object-distinction? Is the zoologist's lac insect indeed the same object as the chemical engineer's lacquer-base? To try to formulate an answer seems to me first of all to involve drawing a distinction between object and thing; that is, object may be a complex idea, with thing and aspect as its elements. Thus to oppose object and aspect is fallacious in a deeper sense than the linguistic agnostics such a Whorf, Benveniste, and (in our own field) de Grolier might call that between thing and aspect.

.35 FURTHER CASES IN LC

The questions above, about distinctions within History, have an interesting answer in the practice of LC: "m.c." C..., Auxiliary Sciences of History, is clearly and admittedly a congeries of several disciplines that, while historical, are not history proper. Their objects are not events (in the sense implied by the placing of a war in D..., E..., or F...) but constituents of events such as coins, physical documents, inscriptions, contents of gravesites. But is the numismatist distinguished by LC as a fully disciplinary specialist, or is the archaeologist? The answer seems to vary from case to case, in this sense: Numismatics (CJ...) is indeed an m.c. insofar as its objects are all collocated in CJ...; but Archaeology (CC...) is not, since the treatment of the archaeology of Peru is a subclass in F... or GN..., subordinated, that is, to the place. (Coins of Peru, on the one hand, are classed in CJ2380-2399, while prehistoric Peruvian pottery goes to F3429.3.P8.) Place as primary division of History seems to emerge as that which unifies the specialism; LC seems (I venture momentarily into explanation) to imply that for some specialists place is so dominant as to override the specialism itself: what matters most is that this is a Greek gravesite, not that it is the incidence of a particular style of *thólos*. The classing of the idea, and hence of documents on it, and hence the expectation of specialist-interest, is determined first of all by the place and only thereafter (if at all) by the archaeologica characterisitics of the object of study. In Numismatics, to repeat, things are different: both specialist- and place-specification is kept within the discipline's self-contained boundaries.

.35+ NOTE 1: ARCHAEOLOGY IN L C

Archaeology in LC is in fact a rather mixed case: gravesites, castles, cooking implements, etc.--almost all archaeological objects--are classed with the place or race that they belong to; but two large and four small classes are provided for treatment of certain archaeological objects. in the two "large" and in one of the "small" cases divisible by place· Bells Crosses; Westhercocks, Fibulae, Stone heaps, cairns, etc., of unknown purpose, and Monograms. Description breaks down in face of such whim.

.35+ NOTE 2: ANTHROPOLOGY IN L C

Another example of the dependence of one specialism upon another can be seen in GN... (in LC), Anthropology, since only generalities, races and ethnic groups (with "purely descriptive treatment of peoples and races" in D.../F...), and prehistoric archaeology are allowed. This case is unusual, though, in that "m.c." G..., with Anthropology at GN..., follows the classes D..., E..., and F..., rather than the usual situation, namely with generalia topics preceding the more concrete divisions of the subject at hand[3]. (The presence of stone heaps of not entirely known purpose here too (GN790-796) as well as at CC700-705 raises questions of consistency just as much as does the distinction "purely descriptive" as applied to races classed in D.../F...: What is the distribution of documents between the two limbs of the bifurcation? Can I be sure I have all on race "Q" from an anthropological point of view when I have finished looking in GN537-686, or must I now try D.../F..., and vice versa? If it is *I* who am uncertain about the purpose of the stone heap, shall I try CC700-705 or GN790-796? Is prehistoric archaeology of a place (GN...) or that place's antiquities (D.../F...) the best locus for such a search?)

.36 SUMMARY OF MAIN CLASS IN L C

What we see here then is either a relatively narrow m.c.:--
History, with its dependencies (Archaeology, Chronology, Anthropology,

---

[3] *Cf.* my paper "Citation Order--Presuppositions, Structure, and Function", pp. 22-32 in *Towards a Theory for UDC; Essays Aimed at Structural Understanding and Operational Improvement* (London: Bingley, 1969).

etc.) and its collateral but tributary classes (Diplomatics, Epigraphy, Heraldry, Genealogy, Numismatics, etc.); or else a very broad m.c.--History taken as wide enough to embrace Archaeology, Anthropology, Diplomatics, Epigraphy, Genealogy, Numismatics, etc. The only difference between these two views seems to me that collateral classes such as Numismatics are demoted by the second view; why not then fully destroy their autonomy by making them nothing more than dependencies?

.37 CASES IN UDC: EDUCATION AND LIBRARIANSHIP

Do we see similar treatments in UDC? A perhaps legitimate vestige of the main-order-symbol definition of m.c.s is constituted in our taking 37... (Education), as an m.c. within 3... (Social Sciences). 37... bears to its notational superordinate 3... one of the relations that obtain between Archaeology and (on the other hand) Numismatics to History in LC. But 3..., Social Sciences is just as much a congeries in UDC as was C..., Auxiliary Sciences of History in LC, so the analogy is lame.

As against those subclasses of 0... in UDC which have form-parallels in (0$\underline{n}$...), there is a subclass that has the appearance of an m.c., namely 02... Librarianship. That is to say, it has within its notational boundaries all specifications--whereas those subclasses of 0... that do have form-parallels in (0$\underline{n}$...) use these latter to become subordinated to any subject throughout the schedules. These are 03... $\cong$(03...), 04... $\cong$(04...), 05... $\cong$(05...), 06... $\cong$(06...), and 08... $\cong$(08...). But these examples are still lame as analogies to the LC manner of making one class dependent on another.

.38 IS THERE A PARALLEL BETWEEN C... (LC) AND 3... (UDC)?

Perhaps this can be seen in similarly oriented disciplines--there may be something about the amorphousness of History that attracts the surgical approaches of amputation (Numismatics) and grafting (Archaeology) --though there would be different classes thus demoted in the two classifications. But examination of the various second-order divisions of 3... does not reveal the sort of dependency we see in LC: 30... (General Sociology), 31... (Statistics),32... (Political Science),33... (Political and Social Economy), 34... (Jurisprudence), 35... (Public Administration), 36... (Social Relief and Welfare), and 39... (Ethnography)--are all wholly self-contained; in no case do they hold only generalities, with specificities ranged under place in 9..., though DC and UDC are looked at askance for not having more closely related (notationally) these parts of an allegedly single super-class.

.38+ NOTE: PARTS OF 3... NOT RELATED TO HISTORY

I do not see 37... (Education), or 38... (Trade), as candidates for dependency upon 9... . In 38... we see a different sort of dependence than hitherto mentioned, that namely between two m.c.s or parts of two m.c.s, here with 65..., Commercial etc. Techniques, etc. But the same split into two of what some might well consider a single m.c. has been performed on other subjects both in LC and in UDC, i.e. Science and Technology. And what in UDC are two m.c.s (Geography and History, 91... and 93.../99...) have been adjoined in LC, in D..., E..., and F... .

.40 SUMMARY COMPARING SEVERAL CLASSIFICATIONS

To summarize up to this point: UDC is made up of autonomous m.c.s, but they are not notationally evident; there are occasional splits into two of what in other classifications constitute a single m.c.; and in the case of 0... (Generalities), 3... (Social Sciences), 5... (Mathematics and Natural Sciences), 6... (Applied Sciences), 7... (The Arts), and 9... (Geography/Biography/History) it is clear that there is some attempt at a provision of super- or summary-classes, though certainly not in a helpfully systematic way. LC, as we have seen, is made up of m.c.s, with tributary quasi-m.-c.s in relative proximity, the whole forming super- or summary-classes--most clearly so in the case of C.../G..., which we might call "Historical Sciences". CC improves on both these classifications by introducing a more helpful over-all order, as does BC as well.

.41 DETAILED SUMMARY OF UDC ALONE: THE DEFINING FUNCTION OF SPECIAL AUXILIARIES

When we look more closely at UDC's m.-c. structure (not comparatively, but of course not discounting the benefit of our previous examinations), we see that there is a notational definition of something, even if not of the m.c.: this is in terms of special auxiliaries applicable to all the (disciplinarily defined) m.c.s within a (notationally defined) m.c. For instance the set of m.c.s 31..., 32... . . . 39... (even including the non-history-related 37... and 38...) all may make use of any special-auxiliary codes provided at 3..., in other words the auxiliary of Persons. The important point to note here is that the range of applicability of such a special auxiliary defines what belongs to a peculiar type of class, namely "the-class-to-which-such-a-subordinate-is-appropriate". That the 3-05 special auxiliary for Persons has become a common auxiliary -05 is a severe negative blow to the previously apparent unity of 3..., however generally useful. The technical and stylistic special auxiliaries given at 7... serve to draw the various arts 71..., 72... etc. together, but variable use of them shows a loosening of the supposed m.-c.- or super-class unity: in 78... (Music) there is a development of the 7.08 special auxiliary that is reserved to 78... alone; thus the whole range of 7.01/.09 special auxiliaries is not necessarily unifying if it merely serves as an over-generalized congeries of special auxiliaries applicable only to second-order m.c.s. The fact is, though, that 78... is the only such second-order m.c. that has such a specially developed sector of 7.01/.09. But the fact of this special development seems to me of capital importance: it tells me that Music is not as closely related to Architecture as is Sculpture. This of course is no news, and it should better have been stated the other way 'round. But since some of the characteristic concepts subordinate to the visual arts are also appropriately subordinate to Music, it seems a sensible decision of Dewey and his successors to have put the latter into a super-class with the former.

.41+ NOTE: MAIN-CLASS STATUS OF 79... (UDC)

Intuition tells us that even more remote than Music from the visual arts is Entertainment etc., 79... . In LC this subject is not in N... at all, but rather in G..., P..., etc. Yet it shares the use of the 7.01/.09 special auxiliaries. That this is possible and meaningful indicates its legitimacy, though it does not justify consideration of 7... as an m.c. Indeed, the multiple contents of 79... (Films; Circuses etc.; Theatre; Pastimes; Games of Chance; Sports; Hunting etc.) bespeaks even it (notationally already of the second order!) as a super-class, not as an m.c. If any proof is needed of notation alone as a misleading definer of m.c.s, look at this case.

.42 CONTINUATION: THE DEFINING FUNCTION OF PARALLEL DIVISION

At several subdivisions of 632... we are told to divide as some code in 58... or 59... is divided. 632.6... (Animal Pests) may be divided as is 59... (Zoology), 632.51... (Weeds) may be divided as is 582... (Systematic Botany). (In each case there is also allowance for alphabetical arrangement by name of vermin or weed, but in this case there is the danger of Latin and vernacular names for related things being used together indiscriminately, or of several vernacular names for the same thing being used together, or both.) The presence of such parallel division indicates a similarity between the things so divided.

.43 A WEAK SENSE OF MAIN CLASS AS DEPENDENT ON SHARED SUBORDINATE CONCEPTS

What then of those things (or subjects) that can be divided by (0n...), that is, all codes in the schedule? There is the similarity between them that they belong to a single class: documents. Perhaps not every thing (or subject) can be meaningfully characterized by a place- or time-division, but all those that can do belong to a class that is distinct from the class belonged to by those that cannot. And surely, then, whatever two or more subjects can be divided by the same more-or-less general subordinate concept show a degree of belonging-together-ness not found among whatever subjects cannot be divided by the same subordinate concept.

.50 SUGGESTIONS FOR APPLICATION

But what does this understanding contribute to the improvement of UDC? After all, we have seen that UDC's m.c.s are far more clearly delineated than are LC's, and are somewhat more helpfully grouped into super-classes than are LC's (though surely not so well so as CC's or BC's); what more can be done? I shall confine myself to two fairly concrete suggestions:

.51 APPLICATION TO SCIENCES AND TECHNOLOGIES

UDC is a-tinkering, and even the most entrenched present file-administrators cannot prevent its eventual improvement. Reclassification after all is a duty of every librarian and information-center manager, even if his classification does not experience periodic new editions (DC, CC), but is only in constant flux (LC, UDC); and it is all the more the case that a classification must attempt to keep up to date so that new users will not be taking on a product that is, so to say, secondhand.

Despite D.J. de Solla Price's legitimate insight into the disparate origins and often-misunderstood relations between the sciences and the technologies, it is possible to see a relationship between pairs of classes in each super-class, relationships stronger than those between many classes of the sciences or of the technologies. What do 61... and 63... (Medicine and Agriculture) on the one hand have in common with 66... (Chemical Engineering) that can compare with their bond to 57.../59 (Biology)? If they can justifiably remain in separate classes is it not due largely to the difference between their special auxiliaries, derived from the "pure" nature of the 5...-classes as against the "applied" nature of the 6...-ones? (Systematic taxonomy, for instance, is unhelpful in arranging pests and vermin, since some grain-pests are of species closely related to the crops they infest.)

Consideration should therefore be given to undertaking analysis of a reorganization of the total domain of science, intending to retain in 5.../6... only those classes which can justify their presence by the central test of hierarchical warrant (to be explicated shortly), and grouping the pairs "science$_{\underline{a}}$/technology$_{\underline{a}}$", "science$_{\underline{b}}$/

technology$_b$" in terms of their sharing of common concepts, for which a strong test is the use of parallel divisions and shared special auxiliaries.

.52 APPLICATION TO "PURE" SCIENCES

If, due to pressure for file-stability, such a revolutionary relocation cannot come about, the sciences can still be advantageously grouped in a manner analogous to that recommended in the integrative-levels theory being applied by the CRG. Whatever class in the domain 5.../6... is most commonly used as a reservoir of subdividing concepts for any other class in the same domain should precede, as being thus operationally defined as more "primitive"; whatever class (if any) in the same domain can be subdivided by all other classes but cannot subdivide them should stand last. Provisionally, the general plan seems likely to emerge as Mathematics - Physics - Chemistry - Life Sciences - Planetary Sciences, each (where appropriate) accompanied by its technology.

.6 THE PRINCIPLE OF HIERARCHICAL WARRANT

The following may not be considered germane to a descriptive treatment of UDC or indeed of any extant classification--it arises not from anything descriptive at all but rather from a categorial insight into the unintended essence of the practice of hierarchical structuration--but its statement is necessary in order to allow a rethinking of the evidences laid out in the preceding. It is the answer not to the question What is an m.c.? but to that What is a class? It is a purely nomological statement, not the statement of a classificationist but of a classification theorist.

Just as Hulme[4] has argued that (as Mills resumes it[5]) "a term is not enumerated unless a literature has appeared on it", I would argue that the hierarchical conjunction of multiple classes at a lower level depends on the fact that documents at that level treat (or are likely to treat) those multiple topics together. The strength of the class-relationship between disease, therapeutic organic substance, and technique of application, as against that between the same substance (no longer considered in its therapeutic aspect) and its isolation, synthesis, and manufacture, can be expectedly verified in the extant literature to justify the existence of two classes, Medicine and Botanical Engineering. The whole multi-dimensional pyramid of knowledge may then form a whole in that discussions of it in a more or less abstract way are possible, but it will articulate itself into super- or summary classes on the basis of a lack of mutual discussion of certain lower-order topics. Only this summarization-tendency and the economization of information storage that follows from it can justify the hierarchical structure evinced in all classification systems.

---

[4] Wyndham E. Hulme, "Principles of Book Classification", Library Association Record, 1911 (pp. 354-359, 389-394, 444-449), 1912 (pp. 174-181, 216-221).

[5] Mills, op.cit., p. 5.

I have not succeeded in answering my own question to my own satisfaction--namely, What is an m.c.?--because I am unsure how to recognize within each of the broadest domains of the hierarchical structure as a whole a similar principle of articulation. Is it derived from the same mode of "rifting" that fairly clearly articulates (according to the Husserlian account in Ideen II) into Nature - Culture - Spirit, or do we need to take refuge in the "weak" sense of main-class structure mentioned above, the testimony of thinking by analogy revealed in sharing of parallel divisions, and of subordinate modes of though manifested in common use of special auxiliaries?[6]

---

[6] I must testify to the stimulus (and, doubtless, the misapplication) given by Farradane's thought that main classes should be the sums of their objects, not the common ground of their distinctions.

.7 SOME FORMALISTIC RECOMMENDATIONS: THE PRINCIPLE OF AUTONOMY

Let me begin again, in a wholly dogmatic manner: To be an m.c. should mean to have a distinctive notation, to relate to a disciplinary specialism alive in the world of thought and/or work, and to share far fewer subordinate concepts with any other m.c. than are shared among the "home" m.c.'s sub-classes. But it also means (and this is, I believe, most central of all, however odd the conclusions it leads to when used as a measure of extant systems of classification) to have within the notational boundaries all the concepts specific to it (it needs to have under its own "governance" all that is necessary to express its own specific concepts, analogous to a country supplying all its own food, raw materials, and educational services). At ML105-106 are Bio-bibliographical Dictionaries, Lexica, [and] Encyclopedias, the first code being for international, the second for national coverage, A-Z. If periods were available, and musical forms, and media of performance, etc., there would bloom out a virtual m.c. right here. It is the fact that not all the concepts specific to this bibliographical form are available here that prevents ML105-106 from being a mini-m.c. A poorly chosen citation order is what could lead to and over such a beckoning precipice; let us avoid such beguilements in UDC.

.8 CONCLUSION

UDC is not in quite such desperate shape, but we need to provide better for super-classes and for consistency within m.c.s. That the mechanism of evaluation of results (the principles of hierarchical warrant and of autonomy) are indigenous I cannot sincerely argue; but improvement does not necessarily come from within an organism in some degree of pain. What I am recommending is not a change of personality, but a freshening of atmosphere, an improvement of diet.

There is still a long way to go.

AN INVERTED UDC
INDEX AS A STEP TOWARD SIMPLIFIED AND EFFECTIVE THESAURUS DEVELOPMENT AND MAINTENANCE

Malcolm RIGBY
National Oceanic and Atmospheric Administration,Rockville, Maryland, USA

During the past decade tremendous amounts of time and effort have gone into the compilation and maintenance of special subject thesauri as growing computerized indexing and retrieval systems bogged down because of growing vocabularies.

It has been shown that new or supplementary approaches to indexing and retrieval are required when the volume of materials in a system or "data bank" increases beyond certain critical limits. Roughly these thresholds occur each time an "order of magnitude" of growth is passed.

The same rough rule applies to growth of vocabularies - closely associated with growth of the data bank itself, for the same psychological and logistic problems are involved in vocabulary control as in recrod control - in other words there is a built-in correlation between size of system and size or nature of the searching vocabulary and complexities due to size of one parallel those of the other.

Table 1 shows the general nature of the curves of growth and types of approach to subject retrieval. Table 2 gives a generalized approach to the problem of vocabulary control as related to size of system:

We will not consider in any detail the first 3 classes (<1000 terms) as they do not require formal or systematic treatment-merely careful listing, some cross-references, and care in adding terms.

Nor will we consider at length the last two classes (>1,000,000 terms) since the problems require a very sophisticated systems approach - i.e. conclusions cannot be drawn from intuition or observed experience, but only by controlled design and testing.

Tbale N° 1

Sime of Collection vs. Subject Retrieval Approach

Scale No.	Log Scale	Number of Documents (average)	Number of Subject Classes	Terms	Type of System	Examples of Collection
1	10	50	5(1/10)	5(1/10)	Random	Office-desk
2	10^2	500	25(1/20)	50(1/10)	Browsing	Home Library
3	10^3	5,000	125(1/40)	500(1/10)	Mnemonic	Departmental Library
4	10^4	50,000	625(1/80)	5,000(1/10)	Abridged Control	Special Library
5	10^5	500,000	3,125(1/160)	50,000(1/10)	Detailed Control	Major Public or Govt.
6	10^6	5,000,000	15,625(1/320)	500,000(1/10)	Multi-Access	Univesity
7	10^7	50,000,000	78,125(1/640)	5,000,000(1/10)	Complex Universal	Worldwide System
8	10^8	500,000,000	400,000(1/1280)	50,000,000(1/10)	Ultimate	Possible Universe

Table N° 2

Vocabulary Size vs. Vocabulary Control

	Vocabulary Size		Characteristics	Type of
Class	Range	Average	Use/Need	Controls
1	(1-10)	(5)	no need	none
2	10-100	50	mnemonic	none
3	100-1000	500	browsing	classif. or
4	1000-10,000	5000	coordinate	thesaurus
5	10,000-100,000	50,000	classif.with index	dual-access
6	100,000-1,000,000	500,000	class, index,or S.H.s	triple-access
7	1-10 million	5 million	universal	multi-access
8	10-100 million	50 million	heterogeneous	all possible

This boils the treatment down to classes 4,5 and 6 - i.e. the range from 1,000 to 1,000,000 terms to be assigned to growing data bases ranging from 10,000 to 10,000,000 records (or documents). This encompasses most live and growing abstracting and indexing services, special libraries, university libraries, and even the main subject collections of the Library of Congress.

It also encompasses the magnitude of most vocabularies, thesauri and general classification schemes or subject-heading lists, where the volume is in the following order of magnitude range:

First we will discuss the characteristics of vocabulary control for classes 4,5 and 6 one by one:

Class 4 (10,000 - 100,000 records with 1,000 - 10,000 terms)
/abridged UDC/

The controls here are manageable either by manual or machine methods. Subject specialists are not required for added terms, though some research is required. One person can easily control the vocabulary. Classification can replace semantic access if language is a problem either because of a multi-lingual body of users, or because words are more difficult to define precisely than class-codes. An abridged classification scheme or a rigidly controlled vocabulary is possible.

The traditional thesaurus tends to replace the coordinate systems (such as "uniterm") as the number of terms passes 5,000, especially if the data base is somewhat heterogeneous. But the fact that a thesaurus is easily developed and controlled at this stage is a concomitant of the fact that it is not really needed except by those unfamiliar with the system.

Class 5 (100,000 - 1 million records - 10,000 - 100,000 terms)
/medium edition of UDC/

As the number of terms passes 10,000 the controls become more and more complicated and difficult to maintain. Machines are essential for efficient vocabulary control. Each term added to a thesaurus has an increasing number of connections, each of which must be examined carefully before decisions are made.The cost of adding, deleting or modifying increases exponentially. A thesaurus of over 20,000 or 25,000 terms is completely uneconomical if not impossible to control or maintain. To add 1,000 terms would cost $ 100,000 and 5,000 terms $1 million. Yet this is the size-range that most major and many minor services find themselves in. The only answer is to get out of the bind of the concept of a rigidly controlled thesaurus, which at best is a simplistic compromise between a vocabulary or subject-heading and a classed system.

What are some of the logical and practical reasons why a thesaurus is no longer efficient in a rapidly growing system?

1) As the vocabulary grows the relationships, especially to the distant relationships, become less meaningful and more cumbersome, yet must be considered if the network is to be maintained.

2) As the vocabulary grows, supplementary aids must be invoked to allow controllers and users to "find their way around the map" - e.g. collected lists of terms, structured lists of terms, concordances, micro-thesauri, etc. all of which are used with the EJC or TEST Thesaurus.

3) A general thesaurus is never detailed enough for any of the specialized disciplines or fields of interest which need 1 or 2 order of magnitude more detail in their own specialty.

4) The compromise to the compromise for a thesaurus, getting too unwieldy and expensive to control, is to have a controlled and an uncontrolled vocabulary - say 5,000 controlled terms and 50,000 - 100,000 loose and uncontrolled or semi-controlled terms.

These "hidden" terms can be searched by machine, but are "second-class sitizens". They include:

- new terms may eventually be included in the thesaurus
- terms representing trade names or often used colloquial terms
- biological species or chemical compounds
- terms from unrelated disciplines which might be significant
- geographical place names
- names of people or insitutions or corporations of subject interest
- adjectival terms which might be coordinated for retrieval

The logical and practical (though not the perfect) solution is to go to a dual system - use both class systems (in detail) and direct entry subject headings. If a class system has over 5 or 10 thousand sub-classes, it needs an alphabetical index of about the same magnitude and detail. If a verbal system, whether a thesaurus or a subject-heading, reaches 5 or 10 thousand entries, it needs a classed structure to allow the assigners, users, and maintainers to know what terms might be found under a given concept to eliminate absolute synonymns or inversions in word order while preserving necessary variations which have significance, and to help in cross-referencing and editing when additions or deletions are made.

So if one must maintain these dual systems for internal use, why not give the user of the system the advantage of an optional approach depending on his habits of mind, the nature of the subject matter or the collection, or whether he wants a few very relevant or all relevant itmes for a subject or a field.

Here is where the machine or computer can be of greast assistance. Since the listing of terms with class numbers or of class numbers with index (A/Z) terms is a simple mechanical operation, even when more than one language is involved, a sorter and tabulator can be used at first, two card decks can be maintained if a little larger system is involved (one in A/Z and one in 0/9 order) and when the system reaches 20,000-50,000 or more entries, a computer can be economically invoked.

Experts may argue over whether or not computers should be used for searching or retrieval by class rather than by natural language - nevertheless no one would argue against the use of computers in maintaining tools (i.e.vocabularies, thesauri, or class schedule if such).

A few examples of one or several language schedules or indexes developed by tabulator and computer for operational use with systems of class 4 or 5 magnitude (50,000-500,000 recrods and 5,000-50,000 terms) are appended.

Class 6 (1 million - 10 million records - 100,000 - 1,000,000 terms)

The main purpose of this paper is to suggest a possible mechanism for extending the dual-access subject retrieval approach to a multi-access approach which would facilitate control of data as well as of vocabularies and class systems of this order of magnitude.

The "data bases" of this magnitude comprise all or large portions of the following Services or Institutions:

1. CNRS (Paris) and VINITI (for 10-15 years accumulation)

2. Library of Congress, Science Museum Library, Bibliotheque National (Paris), Lenin Library, etc. (main collections only)

3. Chemical Abstracts
   Biological Abstracts
   Engineering Index
   Defense Documentation Center and NTIS

These are not quite "universal" but need a universal approach as they grow, with time, to the next order of magnitude (10-100 million).

The "vocabularies" or class systems here are of the general order indicated in the model, i.e.

Library of Congress Subject Headings (> 500,000 terms)
Universal Decimal System and L.C. Classes (>150,000 classes)
English or German Unabridged Dictionaries (>500,000 terms)
Chemical Compounds in CA Register (1 million)
Place Names in Board of Geographic Names Lists (>10 million)
Biological Species according to Smithsonian (>1 million)

So the "roof" for the last 3 bases for terminology are already getting into the class 7 (1-10 million or over) category which must be kept in mind when working with "universal" systems reaching 1-10 million records in a decade or two of cumulative operation.

Methods of extending the dual-access to a multi-access approach

What other methods of access can be applied to a quasi-universal data base to get around the complexity of term control and retrieval? Several suggest themselves:

1) Divided Catalogs or Facets - e.g.

   a) Subject
   b) Chemical
   c) Biological
   d) Geographical
   e) Political
   f) Time
   g) Language

This reduces the number of terms or classes an order of magnitude and makes it easier to manage each. The categories are fairly obvious and not too overlapping such as are the Ranganathan Facets of personality, etc.

2) Coordinate Systems,even using pre-coordinated terms or class numbers to coordinate to achieve a higher level of coordination.

This may involve coordination of any two or more of the types of concepts or facets given in the Divided Catalog "approach", or two or more of any one type or facet, either in fixed, subordinate, or random order.

For instance, one could coordinate lake water (subject) with sulphuric acid (chemical) with crustacea (biology) with Lake Erie (geographical) with Ohio (political) with 1965-69 (time) with English (language) or lake water with pollution with spring floods.

RALAY-SWITCHES
AND RETRIEVAL-DIRECTORIES FOR INTERNATIONAL DATA-BASES AND INFORMATION NETWORKS

Kjell SAMUELSON
Tjänsterförsändelse,Kungl.Tekniska Högskolan,Stockholm,
Sverige

Introduction

The following article might be considered the second part of our previous work (1) on information network switches, that was presented at the FID/CR:CCC:TM:OM seminar in Frankfurt. The topic in essence deals with the feasibility of relay-switches and referral-directories for international information networks and interconnected data-bases (2). Sofar, we have concluded in past studies that a crude ordering system is needed, be it a universal classification or thesauri or merged locator files (3, 4, 5). We might want to answer the question, whether the UDC could serve these purposes and fulfill some of the required functions. Therefore, let us examine the constraints and specify the criteria for a desirable ordering system, say, some classification. The following fourteen issues can be listed.

A desirable classification is/has:

- universal/general
- long life-span
- time-saving search-tool
- discriminator
- perceivable/surveyable
- minimized browse-time
- structure
- MNEMOnic coding
- standard(ized) symbols
- low complexity
- flexibility
- extendability
- maintainability
- defined coupling points

We shall examine each of the above issues and try to arrive at reasonable system design concepts (6).

Universality and Generality

The word universal to some readers might mean worldwide usage and distribution. Others might mean that the contents of the classification should cover the universe of knowledge and be general. Others again might refer to universal as language independent. There might exist several other interpretations, but nevertheless, most of them should be satisfied by the "universal" ordering construct. (7).

Life Span

If universality is commendable, then long life span becomes a necessity. Regardless of this, the use of a crude classification in relay-switches of global information networks do call for a long life-cycle of each component, due to the long development period of the overall design.

Time-saving Search-tool

Older classifications were often intended to serve as procedural rules guiding the placement of documents, materials or goods in allocated boxes or on separate shelves. The roles of some classifications have changed today. The primary function of a referral-directory is to serve as an efficient and time-saving search-tool.

Discriminator

A crude classification for a first entry into the fuzzy and diffuse world mass of information, should have efficient discriminatory power. The referral-directory and relay-switch should both be efficient discriminators. This design aspect would have priority, rather than the philosophical viewpoint or so called logical beauty of taxonomic exactness. Instead, the ordering system would have to be satisfactory in operational terms. This approach is pragmatic and leads to a wider use of action-oriented partitioning, usage tags and indicators of say, technical report, trade brochure, handbook, review, monograph etc. To some extent this feature is provided in the UDC by auxiliaries for form, point of view, language, date and time. Other important features would indicate document length and size, picture, graph, table, author's organisation, level of abstraction, reader forum etc. For instance, information for decision making on environmental quality cuts right across most subject areas and disciplines. Auxiliaries and target-oriented tags would facilitate action-directed retrieval.

Procedural steps of a search strategy could be like Fig.1 :

I Eliminate the majority of unlikely material by cut-off at the main class level, i.e. maximize the refused portion.

II Minimize the dubious, fuzzy, border material at the demarcation line, i.e. the subject headings or entry-level stage.

III Search the somewhat likely portion at the indexing level or term/descriptor categories.

IV Search the most likely rest, carefully with precision and with higher resolution when examining the details say, free-text and strings.

V Search the remaining, less likely rest at the same high level of resolution.

Perceivable and Surveyable

The details of the specialized data-base termed relay-switch, might be imperceivable. However, the mode of interrogation and the referral-directory should be perceivable/surveyable. A quick glance when thumbing through its pages should enable the opening of a speedy search dialog. The abridged UDC edition is fairly surveyable even though the codes might not be perceivable. On the other hand many telephone directories, handbooks and encyclopedias (even with an index) are absolutely imperceivable. These were just some everyday examples.

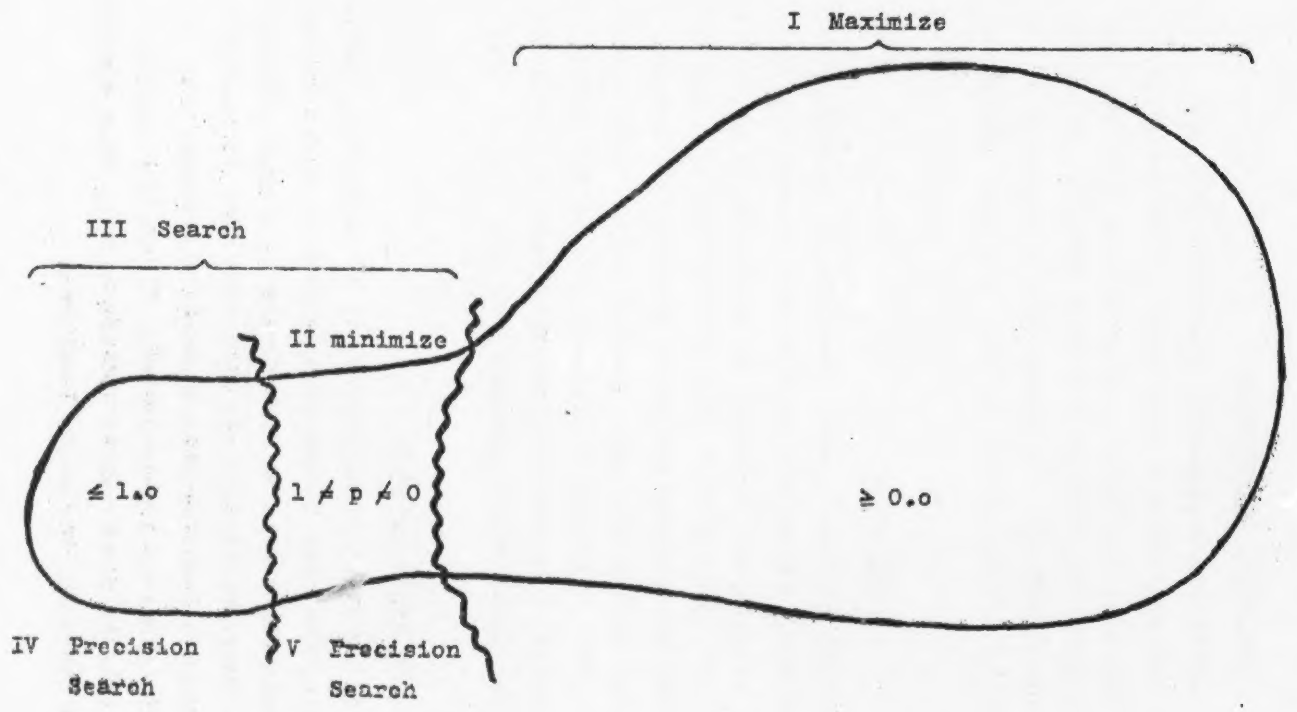

I Maximize
III Search
II minimize
≤ 1.o
1 ≠ p ≠ 0
≥ 0.o
IV Precision Search
V Precision Search

Figure 1. Procedural steps of a crude search strategy at the macro level,

Minimized Browse-time

The need for minimized browse-time applies to both the relay-switch and the referral directory. If the latter is perceivable the time for browsing generally is shorter. The relay-switch, however, requires all those hardware features, that guarantee a fast and easy man/machine interface. We shall not review these aspects, as they have been described elsewhere.

Structure

It is evident that at least a gros structure is necessary for relay-switches and referral-directories even though the fine structure and details can sometimes be neglected at this level of resolution. We shall mention that several isolated data-bases work efficiently on randomly scrambled document collections of limited size. Furthermore many thesauri do not have a structure. Instead, they are merely word lists with SEE and USE references. Even a crude ordering system and inventory index for referral in international information networks must have a structure.

Mnemonic Features

The pocket-size referral-directory has to be managable, convenient and perceivable. Whatever coding or symbolism is meant for use by human interrogators, should be synthesized in a human-factors friendly way. Thus, one should not force the user to handle or memorize numeric codes without a semantic meaning, e.g. zip codes, area codes or log-in numbers. If the code length has to be limited to six or eight characters say, then mnemonics should be available for the major languages.

Standard Symbols

The road toward standard symbols and consequently standardized codes is long and troublesome. It becomes a long-range task. Even though we might not reach its fulfilment, several of the loosely interconnected IR-systems can still remain operative and function in a satisfactory way.

Low Complexity

Complexity as a design concept means approximately the number of interconnections rather than the number of items, components or elements. The referral-directory should have low complexity and should be improved toward simplicity (9). In the relay-switch we may allow a somewhat higher complexity, but its relative complexity should be low, i.e. the ratio between links and components should be low. Any SEE ALSO reference adds to the complexity.

Flexibility

A certain rigity will have to be superimposed in terms of fixed points, forming a stable skeleton and firm structure. This circumstance should not prevent flexibility at most other levels in terms of updating, extension, and maintenance. Dynamic flexibility can be exercised despite the rigidity of an acceptable supra-structure;

Extendability

An ordering system say, a universal classification should have non-limited extendability during the expected life time. This is an estimation which has to be made in operational terms by pure and realistic calculation. Both inward and outward growth should be anticipated.

Maintainability

Maintenance, updating, purging and housekeeping are compulsory features. However, this is a time-consuming task which requires resources in the range of hundreds and thousands of man-years, without exaggeration. These responsibilities are best undertaken by the many concerned expert groups which have been developing their subject-oriented thesauri and special classifications. Thus, the subsystems can be individually maintained by the best suitable expertise. There is little reason for making a big international issue of work which can better be handled regionally and in close physical relationship to the master-bases.

Defined Coupling Points

Many of the issues listed above are related to one major, key issue, namely the necessity of defining the coupling points (8). It appears desirable to continue existing operational IR-systems even in cases where there is some overlap. There is a need for tracing and defining the couplings and linking points. Full compatibility between subsystems might never be reached and iterated conversions are so resource consuming that endless interconvertibility through impossible concordances cannot be justified. However, each particular subsystem can be used separately if the couplings are used as entries and points of reference. Couplings exist at different levels of resolution. For example, you do not trigger an expanded search-routing with multiple "explosive", tree-searches unless you anticipate the amount of output.

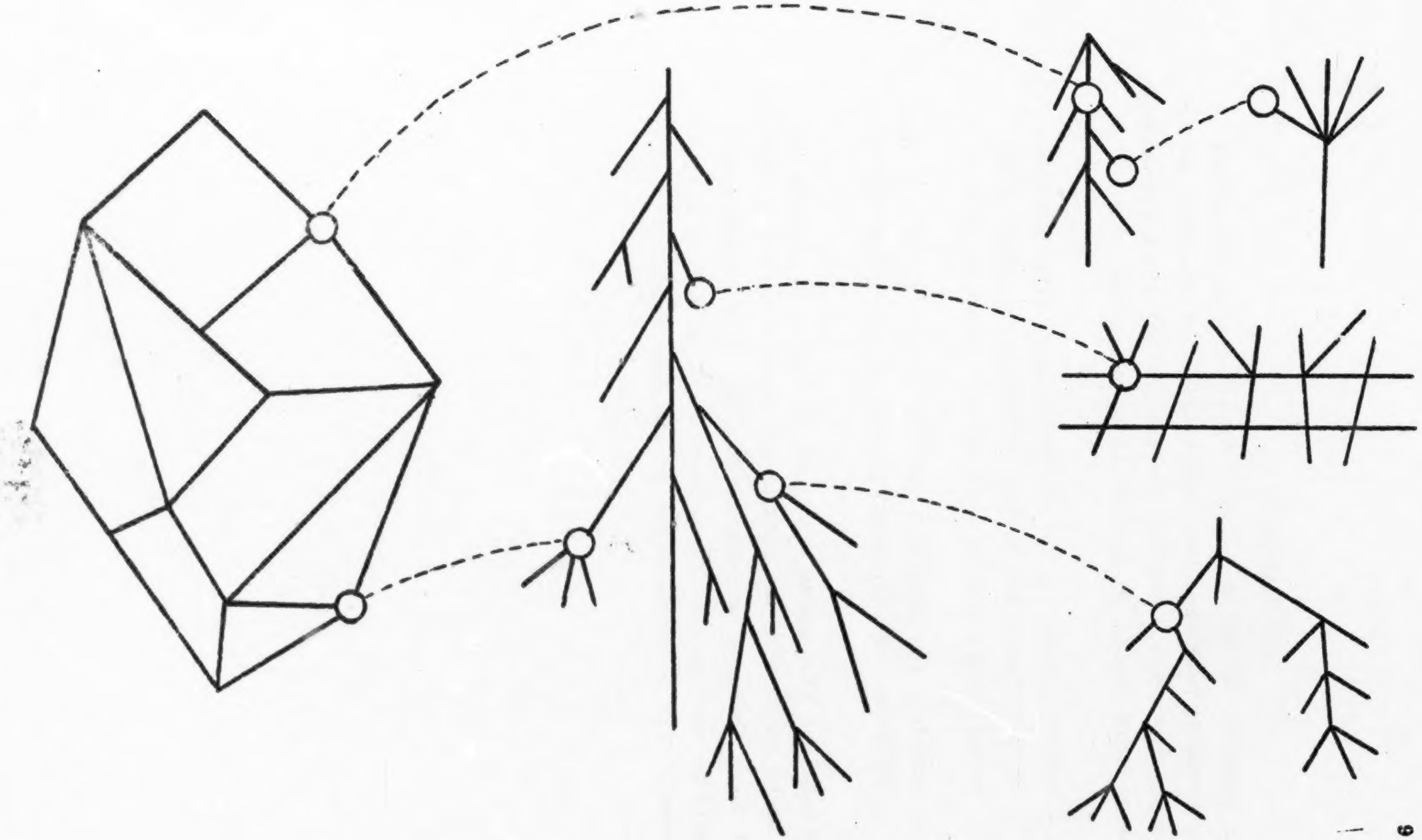


Fig. 2. Interconnection of defined coupling points in classifications and thesauri.

Conclusions

The question of UDC in relation to other classifications and indexing languages has been raised. The fourteen systems design issues above converge toward the following recommendations, which should be applicable to UDC as well:

- update downwards, i.e. standardize internationally at the higher levels and achieve crude, meta-level concordances between well-defined coupling points.
- downdate upwards, i.e. cut the hairy ingrowing bushes, obsolete details, dated phrases and passé conceptions.

The outlined principles are shown in Figure 2. In summary, we might state that the software/brainware for relay-switches and their abridged hardcopy subset, the referral-directory should be studied now. Hardware and equipment have arrived already, with no hints on how to use them intelligently.

References:

1. SAMUELSON, K. Coordination of diversified data-bases in information networks through multi-classificatory concept notation in relay-switches and referral-directories. In: International Federation for Documentation (FID), Central Classification Committee (CCC), Classification Research (CR), Operational Machine Techniques and Systems (OM), Theory of Machine Techniques and Systems (TM) Seminar, 2nd, Frankfurt, 1-5 June 1970. Proceedings.

2. SAMUELSON, K. International information transfer and network communication . In: Cuadra, C. A., ed. Annual review of information science and technology, vol. 6, Encyclopaedia Brittanica, Chicago, Ill., 1971, 269-316.

3. BERUL, L. H. Technical means for national information systems. In: Mikhailov, A. I.; Chernyi, A. I.; Gilyarevskii, R. S., eds. International forum on informatics, VINITI, Moscow, 1969, 149-167.

4. SAMUELSON, K. Systems design concepts for automated international information networks. In: American Society for Information Science Annual Meeting, 32d, San Francisco, 1-4 October 1969. Proceedings, vol. 6: Cooperating information societies, Greenwood, Westport, 1969, 431-435.

5. SAMUELSON, K. Automated international information networks, system design concepts, goal-setting and priorities. Presented at the FID/TM panel at the ASIS meeting, San Fransisco, October 1969. Report. Royal Institute of Technology, Stockholm 1970, 58 p. (IB-ADB 70, No 21).

6. BORKO, H.; SCHUR, H.; AMEY, G. X.; SAMUELSON, K. System analysis, an approach to information, FID/TM tutorial session. FID International Congress, 35th, Buenos Aires, 14-24 September 1970. Report. Royal Institute of Technology, Stockholm 1970. (IB-ADB 70, No 32).

7. SAMUELSON, K. World-wide information networks. In: Conference on Interlibrary Communications and Information networks (CICIN), Airlie House, Warrenton, Va., 28 September - 20 October 1970. Proceedings, Interlibrary communication and information networks. American Library Association, Chicago, Ill., 1970.

8. LANGEFORS, B. Systems for control of organizations. Royal Institute of Technology, Stockholm, 1967.

9. SAMUELSON, K. Management information systems and nervous systems; A comparison of design concepts. In: IEEE Systems, Science & Cybernetics Conference, Philadelphia, Pa., October 1969, Proceedings, 90-98.

SOME REMARKS
CONCERNING THE COMPARATIVE EFFICIENCY OF UDC AND DESCRIPTOR LANGUAGES

Eugeniusz SCIBOR
Central Institute of Scientific,Technical and Economic Information,Warsaw,Poland

1. Preliminary remarks

UDC is a world-wide international system, used by more than 100.000 institutions. Besides descriptor languages are recognized modern indexing languages, which can be effectively used in mechanized information systems. Furthermore, the general opinion is that these languages can well satisfy particular, highly specialized users' interests. Descriptor languages are in some way "fashionable" nowadays. Owing probably to these circumstances the efficiency of UDC and descriptor languages have been very often measured in the frame of experiments on comparative analysis of information retrieval systems. In this paper, results of some experiments and tests concerning the comparative efficiency of UDC and descriptor languages are briefly discussed. These results are supplemented by some conclusions based on documentary practice.

2. Review of experiments concerning the comparative efficiency of UDC and descriptor languages

2.1. TDCK tests

At the 35th Annual Conference of ASLIB the results of an experiment carried out at TDCK /Armed Services Technical Documentation and Information Centre, Netherlands/ were presented by J.A.Schüller /1/. This experiment consisted of two tests concerning the comparative efficiency of UDC and the Uniterm system. During the first test 100, and in the frame of the second

200 queries, had been put to both systems. The following results have been obtained:

Table 1

	Factors	
	Relative recall[1]	Relevance
1st test		
UDC	73 %	12 %
Uniterms	77 %	58 %
2nd test		
UDC	81,5 %	x
Uniterms	99 %	x

In the frame of the second test /the search file utilized during the second test contained only technical reports/, the time of average search by uniterms was much shorter than the time of search according to UDC /1½ minutes against 10 minutes/.

As a result of the experiment many valuable conclusions have been formulated, namely:

1/ The Uniterm system has given slightly better results but UDC has procured many valuable documents not found by the uniterms;

2/ Queries on subjects well defined by UDC are quickly recovered from UDC files but requests for documents on very specific subjects, connected with the specialization of the centre /e.g. certain aspects of specific weapons or fighter-planes/ are answered more quickly and completely by the Uniterm system;

3/ From the view-point of the needs of TDCK the conjoint use of UDC and uniterms seems very useful.

---

[1] "Relative recall" - relation of the number of documents retrieved according to a given system to the total number of retrieved documents.

2.2. Aslib-Cranfield project

Since a voluminous literature on Aslib-Cranfield project exists, we shall only remind the results concerning the comparative efficiency of UDC and the Uniterm system. The recall factor for searches accomplished by UDC was 75,6 %, and for searches performed by means of uniterms - 82 %. The results for the last portion of 6.000 documents showed a more visible superiority of the Uniterm system /Uniterm - 87,3 %, UDC - 76,1 %/. The relevance factor, being equal for all four tested systems, can be neglected in this paper. The time of indexing was shorter when using the Uniterm system; this system has been found to be more efficient and much easier than other indexing languages investigated at Cranfield. The possibility of going up and down the hierarchy without difficulty has been estimated as an important advantage of UDC /2-5/.

2.3. Test performed at the Institute of Ferrous Metallurgy in Gliwice /Poland/

In 1964 a pilot test was performed at the Branch Information Center of the Institute of Ferrous Metallurgy /Instytut Metalurgii Żelaza/ in Gliwice /Poland/. In the frame of this test the comparative efficiency of UDC and of a special system used by the Centre has been evaluated. This system is a mixture of a faceted classification /based on the scheme of American Society for Metals/ with descriptor languages. The system consists of:

1/ Main classes denoted by numbers 1/29.
2/ Auxiliary numbers, denoted by numbers 30/99.
3/ Abbreviations or acronyms, taken from the natural Polish language, e.g. PLE - Electric furnaces /in Polish: Piece elektryczne/.

Example of a full symbol in the system of the Centre:
9,78, NUS - Casting of unkilled steel
where:
9 - Casting
78 - Steel
NUS - Unkilled steel /in Polish: Nieuspokojona stal/

We see that the system used in the Institute of Ferrous Metallurgy was not - in the strict sense - a descriptor language. Nevertheless this system has been taken into account in this paper because it has two features in common with these languages, namely:

1/ The use of keywords /in abbreviated form/, taken from the natural language,
2/ The use of an inverted file, composed of uniterm cards.

The search file consisted of 416 abstracts in the field of metallurgy. Five queries have been put to UDC and to the special system. The recall factor for searches accomplished by UDC was better than for searches perfommed according to the special system. The time of indexing was equal for both systems, but the retrieval time was much shorter in the case of the special system. Short symbols were considered as a great advantage of this system. On the other hand the international character of UDC was considered as an important advantage of this classification /6/.

2.4. Experiment performed by VINITI

A comporative experimental study of the efficiency of UDC and of a descriptor language has been performed in USSR /probably in 1966/ /7/. A search file consisted of about 500 abstracts on radio amplifiers; these abstracts had been published in the "Referativnyj Žurnal". About 30 queries have been put to both systems. The descriptor language has been without grammor. UDC has been used in a way similar to coordinate indexing. The contents of abstracts have been very strictly, word by word, translated from the natural language into UDC numbers. The following results have been obtained:

	Noise factor	Loss factor
UDC	9 %	8 %
Descriptor language	32 %	23 %

We may deduce very easily that the recall factor was 92 % for UDC and 77 % for the descriptor language; 91 % of relevance has been obtained for UDC and 68 % for the descriptors. We see that - in the frame of this experiment - UDC showed a much better efficiency than the descriptor language. The consistent application of the method of coordinate indexing has been evidently the reason of the great efficiency of UDC. We may guess that the failure of descriptors was probably due to an unsufficiently developed vocabulary and to an unprecise indexing. Dr Fomin reported in his paper that an average abstract had been described by about 20 UDC numbers but only by 6-8 descriptors. The author stated that the majority of failures when retrieving documents was caused by inaccurate indexing methods.

2.5. Test carried out at CIINTE

A small test concerning the comparative efficiency of UDC and keywords[1] was carried out at CIINTE in 1969-70 /8/. The search file consisted of 549 abstracts concerning the subject "Information services in general and information activity". Eighteen queries had been put to both systems. The following results have been obtained:

Table 2

Parameters Systems	Recall	Relevance
UDC	37,3	100,0
Keywords	64,0	98,0

Keywords have been found to be more effective than UDC. For sixteen queries better recall has been obtained when using keywords. The retrieval time has been also shorter in the case of keywords.

---

[1] Keywords selected for the computer-printed KWOC index to abstracts published in the abstracting periodical "Przegląd Piśmiennictwa Zagadnień Informacji" /"Review of Literature on Topics in the field of Information Science"/

We must note however that for the purposes of the experiment a very rough UDC indexing, performed in the frame of the everyday activity of CIINTE has been utilized, whereas the indexing by keywords has been done very carefully, in connection with the elaboration of the KWOC index. So we must recognize that the two indexing languages were hardly comparable, Much better results would have been certainly obtained for UDC if a very exact indexing had been accomplished when using this system.

3. Some conclusions concerning the comparative efficiency of UDC and descriptor languages

First, we mast state that there is need for further experiments on the comparative efficiency of UDC and descriptors. Up to now UDC has been mainly compared only with rather simple descriptor languages /uniterms or languages similar to uniterms/. In these languages the vocabulary control has not been accomplished /or accomplished only to a very small degree/. Furthermore, in these languages hierarchical and other paradigmatical relations have not been introduced and no grammar has existed. We know that at present more sophisticated languages, based on thesauri, are mainly used. So in the frame of a future experiment we should compare UDC with a descriptor language based on a thesaurus. All means, occuring in languages which are to be compared, ought to be utilized. We must also avoid some shortcomings /inaccuracy of indexing, too small search file etc./, which have occured in the experiments performed up to now. We intend to carry out in Poland /probably in 1972-73/ an experiment on the comparative efficiency of indexing languages. We want to compare three languages: UDC, a descriptor language based on a thesaurus and /if possible/ - a faceted classification. A methodology of such an experiment has been recently elaborated at CIINTE /9/.

We see therefore that conclusions, which can be drawn out from experiments performed up to now, have only a very limited value. Nevertheless we shall try to formulate some preliminary conclusions and to complete them by some other conclusions, based on documentary practice.

We ought to recognize that the time of retrieving documents by descriptors was shorter in the frame of all experiments[1]. In the frame of three experiments /Aslib-Cranfield, TDCK, CIINTE/ better recall has been obtained when using descriptor languages; in two other cases /experiments performed at the Institute of Ferrous Metallurgy and by VINITI/ UDC has given greater recall. TDCK tests showed a much better relevance when using descriptors; on the other hand in the frame of experiments carried out by VINITI and CIINTE better relevance has been obtained when using UDC.

We see therefore that the problem of comparative efficiency of UDC and descriptor languages cannot be solved on the basis of experiments performed up to now. We are not allowed to state that UDC is more or less efficient than descriptor languages. However, it is obvious that the use of UDC in a way similar to coordinate indexing considerably increases the efficiency of this classification; this has been proved by the results of the Aslib - Cranfield project and the VINITI experiment. On the basis of experiments performed up to now we may suppose that - in general - the efficiency of an indexing language depends more on the way in which this language is used /first of all, the depth and accuracy of indexing/ than on its structure.

The international character of UDC has been stressed by the authors of almost all experiments as a great advantage of this classification. UDC is considered as a system which is very suitable for the systematization of universal collections, consisting of documents concerning various fields of science. There is an opinion that using UDC we can better answer very general queries and carry out retrospective searches concerning rather large, "field-oriented" subjects. Adversely, using descriptors we can probably better answer queries concerning very specific subjects, resulting from "mission-oriented" interests[2]. However, we must realize that the introduction of very general concepts into descriptor languages /these concepts mainly appear in macrothesauri/ will enable answering queries on very large subjects also when using descriptors[1]. On the other hand the consistent use of all devices provided by the UDC structure enables answering queries on many specific subjects when using UDC. It is evident that queries on subjects defined very well by UDC can be quickly answered when using this classification; on the other hand some fields of science and technology /e.g. automation, solid state physics, molecular biology etc./ have not yet been well developed in UDC tables, so we can probably obtain better results using descriptor languages /provided that thesauri or descriptor lists for these fields have been already elaborated/; the possibilities of retrieving documents when using UDC will probably soon improve because schemes for above-mentioned fields of science and technology are under activite preparation. However, we must recognize that UDC revision procedure is rather a slow one and therefore descriptors will probably always have a certain superiority as far as retrieval of information on newly developing topics is concerned.

---

[1] However, we do not know the retrieval time obtained in the frame of the VINITI.

[2] This opinion is based i.a. on the paper of J.A. Schüller / /.

The kind of documents stored in a given library or documentation centre may also influence the choice of an indexing language. We can have good results when books are classified by UDC since subjects of books ave usually more general and wider than subjects of other documents. On the contrary, there is an opinion that descriptors are more suitable for the indexing of such documents as articles and technical reports, which are usually concerned with very narrow and specific subjects and are written from a highly specialized view-point; this has been proved by the second TDCK test, when very good results had been obtained using uniterms to retrieve technical reports. Nevertheless, it seems that we can get quite good results when using UDC provided that a deep indexing is carried out and all means and devices given by the UDC structure are utilized; this has been proved by the VINITI experiment and - to same extent -

---

[1] Up to now thesauri have been mainly elaborated from the view-point of very specific needs and therefore they usually do not contain large concepts.

by the ASLIB - Cranfield project. The use of a systematic classification may be preferable when we want, first of all, to avoid the ambiguities of a system based on natural language; this phenomenon occurs in the case of patent searches and therefore - as far as I know - descriptors are not used in patent offices and other institutions dealing with patent information.

We see that both indexing languages have their advantages and deficiencies. Therefore we should foresee a long coexistence of UDC and descriptors. We have three main possibilities: 1/ the parallel or conjoint use of UDC and descriptors /e.g. UDC supplemented by descriptors when we cannot find appropriate UDC classes to express given terms/ in particular libraries or documentation centres; 2/ the use of UDC with a subject index in the form of thesaurus; 3/ the application of UDC as a roof-classification connecting many specialized descriptor langauges.

REFERENCES

1. Schüller J.A.: Experience with indexing and retrieving by UDC and uniterms. Aslib Proc., 1960, vol. 12, no 11, p. 372-389.

2. Cleverdon C.W.: The Cranfield hypotheses. Libr. Quart., 1965, vol. 35, no 2, p. 121-124.

3. Cleverdon C.: Some aspects of information retrieval. Aslib Proc., 1955, vol. 7, no 3, p. 153-156.

4. Cleverdon C.: The Aslib Cranfield research project on the comparative efficiency of indexing systems. Aslib Proc., 1960, bol. 12, no 12, p. 421-430.

5. Sokolov A.V.: Experimental investigation of the efficiency of information retrieval systems /Aslib Cranfield project/ /In Russian/. Naučno-Techn. Informacija, 1966, no 2, p.20-27.

6. Comparative investigation of the efficiency of decimal classification and alphanumeric indexes in the information work /in Polish/. Report of the Institute of Ferrous Metallurgy no 1145. Gliwice: Instytut Metalurgii Żelaza 1964, pp. 28.

7. Fomin A.A.: On the possibilities offered by UDC and some descriptor retrieval systems - a comparative study. Annals Libr. Sc. a. Docum., 1967, no 4, p. 177-181.

8. Cieszewska T., Choroszuszyna J.: Comparative analysis of the information retrieval system by keywords /KWOC system/ and according to UDC numbers. /In Polish/ Warszawa: CIINTE 1970 /unpublished report/.

9. Ścibor E., Tomasik-Beck J.: Comparative analysis of indexing languages /selected problems/. Warszawa: CIINTE 1970 /in print/ /In Polish/.

10. Ścibor E.: Comparative analysis of the application of UDC and descriptor systems to information retrieval /in French/. Information et Doommentation, 1970, no 2, p. 35-39 /Paper presented at the 2nd French-Polish Symposium on the Modern Methods of Information and Documentation, Paris, 15-17 Oct. 1968/.

ВОЗМОЖНОСТЬ ИСПОЛЬЗОВАНИЯ УДК ДЛЯ КООРДИНАТНОГО ИНДЕКСИРОВАНИЯ

Олгерд УНГУРИАН

Польша, Варшава, Крохална 3м 1310

В истории развития информационно-поисковых языков (ИПЯ), обусловленного требованиями, предъявляемыми все более и более сложными информационными системами, отмечался уже не один крутой поворот.

Все, кто столкнулся с этим вопросом, то ли в силу своих практических занятий, то ли в силу проводимых исследований соглашаются, что первым переломным моментом этой истории была попытка двух бельгийских юристов Ля Фонтэна и Отле создать мировую библиографию (сейчас многие назвали бы это мероприятие мировым реферативным изданием). Для этой библиографии, для ее систематизации с целью последующего поиска, понадобился первый информационно-поисковый язык. Именно язык, который был бы понятен всем заинтересованным во всем мире, независимо от того, на каком языке они говорят. Кто знаком с историей УДК, тот помнит, что по мнению Отле одним из основных достоинств этой системы было то, что в ней используются в качестве элементов алфавита арабские цифры, одинаково понятные на всех континентах. Впервые в истории библиотековедения классификация переставала быть внутренним инструментом одной библиотеки, понятным только посвященным, и используемым для упорядочения книг, а должна была стать международным кодом, облегчающим всем пользователям понимание содержания заиндексированного документа, независимо от языка на котором он составлен.

В те времена еще предполагали, что как научные дисциплины и проблемы, так и научные публикации, отражающие наши знания о действительности, могут быть склассифицированы - иначе говоря, возможно построение строго иерархической объективной классификации наук и использование ее для построения аналогичной классификации научных и технических публикаций.

Предоставляя науковедам решение вопроса о возможности построения объективной, правдивой и стабильной во времени иерархической классификации наук, можно рискнуть высказать предположение, что построение моноиерархической классификации научных публикаций нереально в принципе, хотя бы потому, что наука - динамический процесс, находящийся в непрерывном развитии, причем отмечается как непрерывная интеграция так и дифференциация различных научных дисциплин и направлений, в то время как классификация произведений печати должна обладать прежде всего стабильностью. В пользу высказанного предположения можно еще только добавить, что чем более жесткая иерархия, тем хуже "гостеприимность" классификации.

Одним из первых это понял индийский ученый К.Ранганатан, отказавшийся от попыток построения моноиерархической классификации и предложивший взамен принцип фасетного анализа и фасетной классификации. Это был второй крутой качественный перелом в истории ИПЯ. Интересно отметить, что Ранганатан был первым, который разработал теорию фасетного анализа и фасетной классификации, однако уже Отле ощущал недостатки жесткой иерархии и ввел общие и специальные определители,чем дал толчок к полииерархизации УДК. Дальнейшее развитие УДК полностью подтвердило правильность этой идеи.

Качественная разница между иерархической и фасетной классификациями заключается, однако, не только в том, что последняя объединяет множество иерархических ветвей (фасет), сгруппированных по нескольким категориям, в нескольких десятках равных главных классов, но также и в том, что в противовес традиционным иерархическим библиотечно-библиографическим классификациям, пытающимся найти для главного предмета документа соответствующее место в общей системе, фасетная классификация вообще не рассматривает главный предмет документа (или название его содержания), как единого целого, а членит его на простейшие понятия, чтобы затем их заиндексировать (перевести на слова ИПЯ) и вновь синтетизировать для получения окончательного поискового образа документа.

Одновременно теория классификации Ранганатана, выделив три плана: план понятий, план названий и план нотации,обосновала возможность несовпадения иерархической подчиненности понятий - подчиненности нотации. Иначе говоря, была высказана идея, что соподчинение понятий и названий не обязательно должно найти свое отражение в нотации.

Наконец в пятидесятых годах нашего столетия мы стали свидетелями нового коренного перелома в истории ИПЯ. Сам Тауб, предлагая новый принцип индексирования документов набором терминов, образующих его поисковый образ, не ожидал, пожалуй, что его идея станет толчком, который откроет новый раздел истории ИПЯ. Со времени опубликования системы Унитерм не прошло еще и двадцати лет, а мы уже пережили и эпоху неоправданного беспредельного оптимизма, когда казалось, что с разработкой дескрипторных языков все проблемы индексирования и поиска будут решены, и период больших разочарований и пессимизма, который ныне кажется перешел в трезвый реализм.

Не подлежит сомнению – дескрипторные ИПЯ внесли один принципиально новый и очень существенный положительный элемент в процесс индексирования – они позволили окончательно распрощаться с устаревшей идеей поиска для каждой публикации одного (максимум двух) определенного места в общей иерархии предметов. Документ, заиндексированный с помощью дескрипторного ИПЯ, может быть разыскан исходя из любого дескриптора или любой комбинации дескрипторов, входящих в его поисковый образ.

Однако дескрипторные языки по отношению к иерархическим и фасетным классификациям сделали шаг назад – они отказались от символической нотации, а стали пользоваться лексическими единицами (дескрипторами), заимствованными из натуральных языков. Тем самым авторы дескрипторных языков, сами того не желая и даже не отдавая себе может быть в первое время отчета, отказались от идей международного ИПЯ, стали на путь создания национальных ИПЯ. Более того, все те, кто столкнулся с вопросом перевода тезаурусов, столкнулся с непреодолимыми трудностями построения на другом языке полностью изоморфного тезауруса из дескрипторов, содержание и объем которых были бы те же, что и у исходных дескрипторов.

И это вполне понятно, все лингвисты, переводчики, составители словарей отлично знают, как трудно иногда подобрать эквивалент слова на другом языке, а ведь в ИПЯ вопрос усугубляется, во-первых тем, что дескрипторы выступают по отдельности, вне контекста, разъясняющего их содержание, а по сему должны иметь четко выраженные границы объема; во-вторых, тем, что содержание и объем дескрипторов устанавливается довольно-таки произвольно авторами ИПЯ, исходя из специфических черт данного языка и в зависимости от потребностей и развития данной области знания в данной стране (а как известно в разных странах имеются довольно-таки существенные различия в этом отношении, не говоря уже о принципиальных различиях языков), и, наконец, в-третьих, тем, что дескриптор должен состоять из одного (максимум двух)слова, а тем самым исключается описательный метод перевода дескриптора.

Учитывая вышесказанное, необходимо признать, что внедрение дескрипторных ИПЯ в их существующей форме отнюдь не уменьшает языковых барьеров при обмене информации, а пожалуй даже наоборот-увеличивает. Дескрипторные ИПЯ не только не международные языки, но по всей вероятности даже невозможно построить изоморфные дескрипторные национальные языки. А это уже совсем плохо.

Анализируя все то, что в принципе новое и положительное внесли в развитие ИПЯ каждый из перечисленных переломов, попытаемся прежде всего конкретизировать несколько условий, которым, должен удовлетворять ИПЯ, если он должен стать мировым языком науки и техники, языком, который будет объединять все человечество, помогая распространять сокровища научно-технического прогресса.

1. Учитывая непрерывное развитие науки, интеграцию и дифференциацию отдельных научных дисциплин и направлений, невозможно построить стабильную моноиерархическую схему соподчиненности лексических единиц ИПЯ. Поэтому ИПЯ должен быть построен по принципу полииерархической (фасетной) структуры, в которой будет установлена иерархическая зависимость только внутри отдельных классов.

2. Документ не может индексироваться по принципу выделения главного предмета или названия содержания документа, для которых подбирается затем соответствующий индекс в таблицах. Процесс индексирования должен сводиться к информационно-логическому анализу текста документа, на основании которого определяется набор элементарных наиболее существенных с точки зрения поиска ключевых слов и последующему их переводу на лексические единицы ИПЯ. В плане ИПЯ все эти лексические единицы должны быть равновесомы. Фактическая поисковая весомость должна зависеть только от контекста документа, т.е. от того, сколь полезно и эффективно (с точки зрения потребителя) то или иное слово для целей индексирования и поиска данного документа. Тем самым ИПЯ должен быть пригоден для координатного индексирования.

3. Информационно-поисковый язык должен использовать в качестве элементов своего алфавита отвлеченные общепонятные всем (или почти всем) знаки, так как это происходит в математике или химии. Конечные последовательности этих знаков образуют лексические единицы данного ИПЯ.

Резюмируя можно сказать, что эффективный универсальный международный информационно-поисковый язык должен:

1. быть построен по полииерархическому принципу,
2. быть пригоден для координатного индексирования,
3. состоять из конечного множества лексических единиц, именуемых далее индексами-дескрипторами, каждый из которых состоит из конечной последовательности отвлеченных знаков; четкая граница объема и содержания каждого индекса-дескриптора должна быть строго определена.

Такой язык можно конечно построить заново, но более благоразумно посмотреть, нельзя ли использовать для этой цели какой-либо из уже существующих, хорошо известных и зарекомендовавших себя с положительной стороны, ИПЯ.Очевидно вполне для этой цели пригодна всем нам хорошо известная Универсальная Десятичная Классификация.

В настоящее время никто, пожалуй, не рискнет назвать схему УДК моноиерархической. Всё расширяющийся объем общих и специальных определителей уже давно превратил схему УДК в полииерархическую. Очевидно недалек уже тот день, когда ЦКК/МФД вынужден будет пересмотреть деление Универсум на основные классы на первой ступени и отказаться в этом случае от принципа деления на 10 классов (я хочу подчеркнуть здесь, что ратую за отказ от десятичного деления только на первой ступени, только при выделении главных классов; на всех остальных ступенях деления этот принцип обязательно должен быть сохранен).

Иначе говоря, надо подумать над возможностью определения главных классов, равнозначных между собой, независимо от их нотационной подчиненности в настоящее время. Попробую пояснить сказанное примером: вряд ли целесообразно сохранять, например, индексы 6; 67; 68 и тому подобные, так как, в силу их дальнейшего деления на совершенно отличные по тематике классы, нет такого документа, который можно бы было заиндексировать каким-либо из этих индексов. Одновременно создается какая-то видимость подчинения, например, индекса класса 681.3 "Математические машины" индексам 681, 68 и 6, которые, собственно говоря, неизвестно что обозначают. Поэтому предлагается определить ряд главных равнозначных классов, среди которых окажутся очевидно:...621,621.3,621.38, 621.39,....,669,...,681.3...и т.д.. В дальнейшем очевидно можно будет обозначить эти классы буквенными индексами (например, строчными буквами латинского алфавита, при этом исчезла бы полностью из плана нотации видимость фиктивной иерархической подчиненности и сократилось бы число знаков в индексе.

Рассмотрим это предложение на каком-либо примере: предположим, что класс 621.38 "Электроника" признан главным классом и что ему присвоен индекс в виде буквы R, в этом случае рубрика УДК
621.382.3 Полупроводниковые триоды (транзисторы)
получит соответственно индекс R23.

Соответственно, в главном классе 621.3 "Электротехника", предположим обозначенным индексом Р, освобождается класс Р8(бывший 621.38), который через десять лет может быть использован для нужд "Электротехники".

Однако даже существующая в настоящее время структура УДК с общими и специальными определителями позволяет причислить ее к числу полииерархических, а тем самым можно считать, что УДК отвечает первому из вышеперечисленных требований.

Второе и третье из вышеперечисленных требований тесно между собой переплетаются и поэтому будем их рассматривать вместе. Прежде всего, следует отметить, что УДК использует в качестве элементов своего алфавита отвлеченные знаки - арабские цифры и другие математические и синтаксические знаки (опознавательные знаки). Лексические единицы УДК - это конечная последовательность отвлеченных букв. Четкая граница объема и содержания каждого индекса строго и адекватно определены почти на всех языках мира. Здесь не отмечаются трудности перевода дескрипторов, так как адекватный перевод определения индексов (рубрик УДК) может быть представлен в виде подробного описания, а не одного слова. Тем самым нерешенным остается лишь вопрос ограниченности числа индексов-дескрипторов и техники координатного индексирования с помощью этих индексов-дескрипторов.

Для того, чтобы ограничить число индексов-дескрипторов необходимо построить некое подобие тезауруса на базе таблиц УДК, который далее будем называть тезаурусом УДК. Аналогично как при построении тезауруса из множества слов и выражений натурального языка выбираются дескрипторы, так и в нашем случае, из множества индексов УДК (простых, сложных и составных) необходимо отобрать те, которые будут считаться индексами-дескрипторами. Все эти индексы-дескрипторы и только они могут использоваться для индексирования литературы. Все остальные индексы УДК образуют множество индексов-нонскрипторов, которые нельзя употреблять для индексирования литературы. Для перехода от нонскрипторов к соответствующим им дескрипторам вместо обычно используемой в тезаурусах ссылки можно было использовать правило, что применяется ближайший вышестоящий в иерархии УДК обобщающий индекс-дескриптор.

Очевидно на первом этапе построения тезауруса-УДК целесообразно использовать уже проверенные практикой тезаурусы ИПЯ дескрипторных типов для отдельных отраслей знаний. Итак, например, в области химии можно использовать "Словарь дескрипторов", изданный в СССР, в области техники и машиностроения - тезаурус Объединенного совета инженеров, и т.д Работа по подготовке тезауруса УДК заключалась бы в определении индексов для дескрипторов основных ветвей этих тезаурусов. При этом обнаружилось бы много недостатков как в УДК, так и в используемых тезаурусах, что, в свою очередь, способствовало бы их усовершенствованию. По своему внешнему виду тезаурус УДК будет подобен существующим таблицам, отличие заключалось бы лишь в том, что индексы-дескрипторы должны выделяться каким-то образом, например, они могли бы быть напечатаны иным шрифтом, чем индексы-нонскрипторы.

Последний вопрос, который необходимо рассмотреть - это техника координатного индексирования с помощью индексов УДК. Предлагаемая методика индексирования сводится к выбору на основании информационно-логического анализа содержания документа наиболее существенных ключевых слов, достаточно полно отражающих содержание индексируемого документа, причем, ключевые слова не обязательно должны присуствовать в тексте - лицо, проводящее анализ, может использовать с этой целью любое слово, независимо от того, встречается ли оно в тексте или нет, единственный критерий выбора - это поисковая пригодность этого слова.

Второй этап индексирования - перевод всех выбранных ключевых слов на лексические единицы ИПЯ, т.е в настоящее время на индексы УДК, а в будущем, при наличии тезауруса УДК на индексы-дескрипторы. Набор индексов дескрипторов (индексов УДК) образует поисковый образ документа. Поиск - механизированный или же инверсионный (аналогично поиску по Унитерм-картам) по любому индексу или же по любой их комбинации. Возможно использование обычной картотеки с введением нескольких вторичных документов под каждый индекс, однако это очень увеличивает объем картотеки и лишает систему преимуществ координатного индексирования.

В заключении рассмотрим на конкретных примерах предлагаемый метод координатного индексирования с помощью индексов УДК.

1. Статья в журнале: Злотников Г.Г,Зайченко И.В., Янимяги З.К.: Новое оборудование для обработки информации на Весенней Лейпцигской ярмарке 1970 г."Научно-техническая информация. Серия 1. Организация и методика информационной работы" Науч.-техн.сб. (ВИНИТИ).М.,1970,№ 12, 23-24
Приводится описание некоторых экспонатов Весенней Лейпцигской ярмарки 1970 г., в частности данные о новых фотонаборных машинах и аппарате для изготовления дубликатов диамикрокарт на диазопленке

Статья получила по"традиционному"методу классифицирования индекс:
061.43(430.2-35.7)"1970" - Специализированная выставка в Лейпциге в 1970 году
При координатном индексировании набор ключевых слов и им эквивалентных индексов УДК будет следующий:

Репрография	681.6
Оборудование	.002.5
Фотонабор	681.616.7
Диамикрокарты	778.14
Диазокопирование	773.71
Читальные аппараты	778.27
Научно-техническая информация	002
Выставка специализированная	061.43
Лейпциг	(430.2-35.7)
1970 год	"1970"

Учитывая правила построения индексов УДК, поисковый образ нашего документа будет состоять из набора 7 одинаковесомых индексов:
681.6.002.5; 681.616.7; 778.14; 773.71; 778.27; 002.002.5 и 061.43(430.2-35.7)"1970"

2. Книга: Липявкин А.Ф., Савченко И.П.: Архитектура городских зданий и сооружений; уч.пособие для строительных вузов,Высшая Школа,М.,1970,стр.446
Обзор всеобщей истории архитектуры и градостроительства, рассмотрены современные конструкции гражданских зданий и сооружений, а также основные конструктивные схемы промышленных зданий.
Дан обзор объемно-планировочных решений гражданских зданий массового строительства. Более подробно рассмотрены архитектурно-планировочные решения коммунальных зданий и сооружений.

Книга получила по "традиционному"методу классифицирования индексы:
721 Здания в целом. Проектирование. Планировка зданий
624.02 Части зданий и сооружений.

При координатном индексировании набор ключевых слов и им эквивалентных индексов УДК будет следующий

Архитектура	72
История (архитектуры)	72.03
Проектирование (архитектура)	72.011
Сборное строительство	72.002.2
Части зданий	69.02
Индустриализация (строительства)	69.057
Градостроительство	711.4
Общественное строительство	725
Культурное строительство	727
Промышленное строительство	725.4
Жилищное строительство	728
Учебник ВУЗ	(075.8)

Учитывая правила построения индексов УДК, поисковый образ нашего документа будет состоять из набора 11 одинакововесомых индексов: 72(075.8); 72.03; 72.011; 72.002.2; 69.02; 69.057; 711.4; 725; 725.4; 727 и 728.

Библиография

1. Foskett A.C.: The subject approach to information, London, 1969
2. Taube M.,Wachtel I.S.:The logical structure of co-ordination indexing, American Document. 1953, Nr.2, 67-68
3. Vickery B.C.: Techniques modernes de documentation, Paris, Dunod,1962
4. Sayers B.: Manual of Classification,fourth ed. London,1967
5. Ranganathan S.R.:Prolegomena to Library Classification, third ed. London 1967
6. Ungurian O.:Wprowadzenie do UKD (FID 409), second ed. Warszawa,197o
7. Ungurian O.:Aufbau eines alphabetischen Sachregister zur DK, das den an Thesauri gestellten Anforderungen und ihren Begrenzungen entspricht, RGW-Symposium:"Entwicklung von Deskriptorsystemen und ihre Nutzung beim Wiederauffunden von Informationen, Berlin 28.6-1.7.1966, ZIID, Berlin 1966, Teil 2, seiten 118-131

THE AR COMPLEX

ADAPTED SYSTEMS USED TOGETHER WITH A COMMON REFERENCE SYSTEM

Ejnar WÃHLIN
Independent Research and Consultant Activity,Stocholm,Sverige

A. Introduction

With the ideas presented here I shall try to indroduce a new theory of classification systems with the aim of solving two different conflicts, which both, in different ways, are hampering a sound development and obstruct the possibility of world-wide solutions.

The first of these coniflicts is experienced by anybody who wants to construct a classification system. Two different steering principles are them fighting each other, the desire for a logical order of concepts (subjects) and the desire that the system shall correspond to the customer's wish to gather subjects relating to a given field of interest. This conflict is easily discernible in the structure of UDC, where the latter requirement is predominant both in the technological and the social areas. This has the result that many important concepts will be found in two or more places.

The second conflict arises when one has to decide what kind of classification system one should use, a universal or a special system. The universal system has the advantage of representing the whole field of knowledge and all concepts, the specialized system can more easily be adapted to the consumer's need. These two conflicts are in a way the same - between logical order and practical adaption.

One way to settle this problem is to use two systems in combination,one representing logical order, the other practical adaption. The concepts could be arranged in a universal system or a general "reference system", the structure of which is determined by logical - semantical rules only, without any special trend to adaption. The consumer's requirement could be satisfied by a separate adapted system lying over the reference system. The adapted systems should not go very deep in detail since the necessary specifications are made with help of the reference system.

We should thus have an unlimited set of adapted systems all using a common reference system for specification. This is called the AR-complex (A = Adaption, R = Reference).

This idea has been studied rather thoroughly by the author in two different applications, one for product (commodity) classification, and the other for document classification in the building field.

The following basic description of and motivation for the AR-complex is here given by quotations from the author's reports on these two studies (page 2-15).

B. Product Classification

Product classification for industry and trade was studied in 1969 with the aid of a special research grant. The following description is an exerpt from a report on this research.(8) Some smaller alterations have been made in the text and the classification numbers are adjusted to the current edition of the systems, which have undergone some minor changes.

PRODUCT CLASSIFICATION ACCORDING PC-US (Quotation)

PC signifies Product Classification. This is a system specially intended for classification of products (goods, commodities).

US signifies Universal Classification System.This is a universal classification system for documentation purposes, a general concept system. (see also page 12-13)

The reason for using two systems in combination purposes will not give sufficient guidance for logical placement of all products, and even if all products could be placed, they will not come together in groups convenient for industry and trade. If you want a good survey of all products you have to make a special product system. This should be designed for classifying the products themselves,not for knowledge on products. However,the necessary specifications according to origin in nature, material, form, function, method of treatment etc., demand a general concept system from which these specifications can be taken. The universal system is used on a lower level for this purpose.

PC follows the flow from production to consumption. Generally it only goes down two, three of four steps in the hierarchy, that is 2-4 digits according to the decimal notation principle.

Primary groups in PC: (Cf table I page 6)

Group 1 Natural materials contains all materials (goods) taken directly from nature, including materials which have been lightly treated, i.e. undergone a treatment which does not essentially change the character of the product (e.g.cleaning).

Group 2 Intermediate materials (industry intermediates) are products which have been subject to certain radical treatment but are not ready to be used for direct manufacture of articles or for consumption. This group represents raw materials for further technical processing.

Group 3 Structural products represent ready processed raw material for production of articles, machines, buildings etc. To a large extent the products have a certain geometric form (tiles,pipes etc.), but the group will also include certain formless products (e.g. dyestuffs, glue, etc.).

Group 4 Consumption materials are products ready for consumption which serve their purpose by just being consumed (e.g. fuels, food-stuffs, etc.)

Group 5 Articles. things, devices (utility goods). This group includes all products (articles, commodities, devices) with a design espicialy adapted to a given function. These products are more complicated than the structural products. Contrary to group 4 a certain durability is desirable here, even in the lifetime should be short.

Group 6 Buildings and plants[1] differ from earlier mentioned groups by being non-transportable. They are not considered as products in everyday speech.

Group 8 Energy is included here, since energy (electricity, heat) is a market product.

Group 9 includes products composed of parts from several groups (1-6), and waste.

The classification keeps hold of the fact that the aim of the industry to a great deal is to produce new and more complicated products from elementary material. The substance in a consumption product has usually passed through several stages of processing. In certain cases however a product may go unchanged from nature to comsuption (e.g. vegetables, fruits) and in this case it will be defined by double numbers. Which number to be used depends of the purpose. In the retail trade

---

1) e.g. industrial plants

the consumption number is used, in agriculture the number from the natural products groups. Reciprocal references can be arranged.

Subdivision of the primary groups

In groups 1 - 3 the material (substance) is the most obvious principles of subdivision. A certain uniformity regarding the order of the secondary groups has been attained and thus the sub-groups represent mainly the following original materials:

1- Vegetable	4 - Metallic
2- Animal	5 - Stone and earth
3- Fossile	6 - Chemical

In class 4 - 5, where the division is made according to the function of the products, this principle can not be continued but it emerges occasionally. Thus vegetable materials dominates in 41 and 51, chemicals in 46. See table I.

Principle of notations

We want a system where it is possible to give every product a logical determined position, even products which were not considered when the system was constructed,

The notation is generally composed of one PC-number (N) and one or more US-numbers (a, b etc.). A total notation can be written

N- a:b (or N- a:b:c...)

The numbers "a" and "b" are separated with a colon sign. In the US-number you have generally a point after three digits but this point is now substituted by an apostrophe which has only mnemotechnical character.

(It has been considered to exchange the numbers of-4 Metallic and-5 Stone and earth)

TABLE I. THE TWO-DIMENSIONAL STRUCTURE IN PC

Please note the horizontal structure involving certain uniformity in this direction, by which a two-dimensional pattern has been obtained. In a certain degree the second digit indicates:

-1 Vegetable
-2 Animal
-3 Fossile
-4 Metallic
-5 Stone and earth
-6 Chemical

1 Natural materials	2 Intermediate materials	3 Structural products	4 Consumption materials	5 Articles. Things Devices
11 Vegetable natural materials	21 Vegetable intermediate materials	31 Vegetable structural products	41 Food-stuffs	51 Clothes and personal belongings
12 Animal natural materials	22 Animal intermediate materials	32 Animal structural products	42 Stimulants	52 Furniture. Room and building appliances
13 Fossile natural materials	23 Fossile intermediate materials	33 Fossile structural products	43 Fuels	53 Household articles Tools. Sport articles
14 Ores	24 Metals	34 Metallic structural products	44 Explosives	54 Vehicles. Transport equipment
15 Rocks and soils	25 Stone and clay intermediate materials	35 Stone and clay structural products	45 Fertilizers	55 Machines. Mechanical apparatus. Weapons
16 Natural chemicals (minerals)	26 Inorganic industrial chemicals	36 Chemically active structural products	46 Consumption chemicals	56 Physical and chemical apparatus. Instruments
17 Water	27 Organic industrial chemicals	37 Textiles	47 Hygiene products. Cosmetics	57 Graphicalproducts Documents
18 Air	28 Syntetic industrial chemicals	38 Synthetic structural products	48 Drugs. Pharmaceutical preparations	58 Cultural articles
19 Other	29 Other	39 Other	49 Other	59 Other

The apostrophe may be eliminated, e.g. if we have only four digits. Example:

```
     Apple (fruit) is written
                                    1)
            112 - 622'6 : 621'5761
     plant pro-  | fruit       |
     ducts for --  (bot.)      |_____apple tree (bot).
     food
```

Here 622 is "Parts of plants" and 622'6 is "Fruit". 621 means Plant species and 621'5761 is a botanic number for Apple tree. If all product in this group are classified first by plant part and then by plant species, the digits 622 and 621 can be eliminated and you will get a shorter "practical number" (where colon is substituted by point):

Apple = 112 - 6.5761   Maize grain = 112 - 7.331'31

In the field of chemistry the apostrophe has a quite special function, that is to differentiate between an element and compounds containing this element in the negative ion. Example:

```
The mineral Magnesium Silicate is written
   16 -  23'42 :  2322 --
    |    |              |
 Minerals |__silicate    |___ magne
```

23'42 = Silicate. (2342=Silicon) 2322(Magneisum)
In US Magnesium is 23.22 and here (a special case) the point must not be eliminated. In PC-US it can be written without a point.

If all minerals in group 16 have chemical notation, the digits 23 which in US means chemistry, can be eliminated and you get the "practical number" 16-42.22.

---

1) All classification numbers after hyphen accord. to US-1

The above mentioned notations apply to minerals consisting of two components. In more complicated cases the notation principles may be modified. It is also necessary to differentiate between two separate forms of the same chemical compound. An example is Magnesium Silicate that exists both as talc and as asbestos. These products can be identified separately by introducing supplementary numbers, e.g. with two digits,,and arranging vacancies which make it possible to introduce new products. Example:

16-42.22.01 = talc
16-42.22.05 = asbestos

Within the retail trade the assortment is often limited compared with that offered by nature or industry. Then short "trade-numbers" may be substituted for the US-numbers e.g. for food-stuffs. Thanks to vacancies every new product may find a place of its own.This is the same principle as in the case talc - asbestos above.

Some examples of coding of structural products are given in the appendix and coding of a machine type is shown in fig. 2.

As another example let us take an armchair. In US furniture etc. is 464 and armchair 464'15. In PC furniture can be supposed to be 521. An armchair of wood can be coded as

521 - 464'15 : 2611 or shortened 521 - 15. 2611

Flexibility

Within a PC-group specified according to the model N- a:b, where "a" and "b" are complete US-numbers it is possible to use different principles of subdivision by exchanging the positions of the US-numbers:

N- a:b = N- b:a

If in one case it is desired to collect all metal pipes, in another case all form products made by the same metal, then this desire may be satisfied and we can write either:

Metallic structural products - pipes : copper= 34-2976:2521 or
Metallic structural products - copper: pipes = 34-2521:2976

Special supplementary numbers

For detailed registration of products it is necessary to express method of treatment, kind of package, sizes etc. This may be solved in different ways with the use of series of supplementary numbers designed for the special branches. Example:

Vacuum dried fine grained coffee in plastic bags

```
41-275=141.  28.   .35   .47
 |   |   |    |     |     └plastic       (4= packing material)
 |   |   |    |     └──── bag            (3= type of packing)
 |   |   |    └───── vacuum dried        (2= method of treatment)
 |   |   └──── fine grained              (1= consistency, form)
 |   └ coffee (shortening of 682'75 in US)
 └── food-stuffs (PC number)
```

The supplementary numbers is here separated from the PC-US-number 41.75 by a double hypen or another symbol.

If the positions are determined the number can be written
41-275=41.8.5.7

Connection with the documentation system

The PC-US numbers are easily transformed to US-numbers which are valid by document classifications according to US. Generally the PC-number can simply be removed. Sometimes a transformation may be necessary, but this question is not sufficiently investigated. By mechanised documentation it is possible to extract all products containing a certain number within colon or points. Thus corn, meal and flakes of maize all contain the digits 33131 (=maize) within points and thus all maize products can be collected.

Testing of PC

The system has been tested by inserting all products in the Brussels Nomenclature in the two digit structure of PC.Within the groups 1 - 4 more detailed trials have been made, also with products from the lists of different manufacturing companies and sectors of trade. Vegetable materials and food-stuffs have been specially investigated. The experience gained so far indicates that construction of a total system of products in this way is possible.

Examples of classification according to PC-US are in Appendix.

(End of quotation of the report on product classification).

C. Building Classification

While in the case of product classification the background for introducing the AR-idea (PC-US) was the desire to combine a main order adapted to products with the access to a universal set of concepts for specification of the products, the actual cause in the building trade was the different opinion of whether a universal or a special system should be used for registration of building research projects and statistics for research costs. A contributing factor was the desire for a certain conformity with the main division already used by the National Council of Building Research in Sweden, as this (very coarse)division had proved to be a suitable division at the highest level.

The method of using two systems in combination in the building field was first presented in a report to the CIB Working Conference, Dec. 1969.(9) The following quotation from this report goes a little outside the subject of the AR-complex and gives information on a proposed method for indexes.

Universal and special systems

When the classifiaction of documents in the building trade is discussed, the question often arises whether to use either a universal system or a special system (a building system), or to have one system of each kind for separate use. The documents relating to discussion within CIB seem to suggest a preference for a special system.

Building research, however, relates to numerous problems that lie outside the building sector proper, e.g. the properties of clays, wind, conditions, sociological conditions etc., and a universal system offers the advantage, in principle, that it provides room for everything. In a universal system, the information is divided according to an overall view, which is determined in some way by the connections between different fields of knowledge. Naturally, in this way one does not obtain a sequence that relates directly to the building process.

Thus there are strong arguments in favour of both a building system and a universal system. It seems to me that the problem could be solved by using a building system at the highest level, and the universal system below that for the necessary specification.
There are signs that a combination of two systems, one turned towards the users field of activity and the other turned towards the store of information, is a method that may have some prospect of success. A need for two different systems is also indicated by the interest expressed in certain circles within the building industry to use a special system (SfB) in combination with a universal system (UDC). This development would be promoted if the question of both the structures of the system and the manner of their combination were made the subject of fundamental studies.

The proposed building system, called AZ, has been subjected to some minor alterations in relation to the structure given in the CIB report. Here is the current main structure:

A. General
B. Materials and products
C. Safety. Protection. Hygiene
D. Structure. Building elements

E. Installations
F. Construction. Maintenance. Economy
G. Building work (labour)
H. Rooms and buildings
I. Engineering works (Civil engineering plants)
K. Community planning
M. The building trade and society
Z. Other

Example of subdivision:

C. Safety. Protection. Hygiene

1. Statics. Strength of materials
2. Geostatics. Geotechnics
3. Acoustics
4. Outer climate
5. Indoor climate. Heat insulation
6. Lighting
7. Fire protection
8. Damages, general
9. Other

Continued quotation:

A proposal for a new universal system (1) (2) (3) (4)

The universal system used here is referred to as US, although it has previously gone under the designation TUS (Technical Universal System). It has been worked out mainly with the support of grants from the Swedish Council for Building Research. The system has been developed in two versions, US-1 and US-2, of which the latter is closer to UDC, while the former seems to be suitable for specialized technical documentation. This version is therefore used in the present context. US is based on simple concepts, rather than traditional disciplines, and technical applications of basic scientific concepts are brought together with these concepts. This facilitates the classification of documents, and also the incorporation of key-words in the system groups, thus bringing key-word indexing and classification closer together. An essential difference between UDC and US-1 is that US-1 contains a materials and an energy class (Classes 2 and 3), and this constitutes in many ways the character of the system. As a rule, considerably shorter numbers are achieved than with UDC.

An important feature of US is that it relates to simple concepts; complex concepts being formed by combinations of simple ones e.g.

Copper-tube=Tube : Copper or Copper : Tube

In US-1 Copper-tube is coded as 2521 : 2976, which is not an 8-digit number but a combination of two independent 4-digit numbers.
Such designations for materials nad products can be used also for costing and similar practical purposes, perhaps in simplified form. They could be valid in all fields of technology and are not limited to the current building materials market.

Document indexing with combination of numbers and key-words

This can be illustrated by an example relating to a Swedish building research project:

> "Development of more suitbale sound-insulating doors for hospitals". Reg.No 68-0415.

The document can be indexed with three key-word lines with classified key-words.

		Reg.No.
C3 Sound insulation of 334'5	Doors for hospitals 451'75:454'41	68-0415
D75 Doors 451'75	Sound insulation of 334'5	68-0415
H41 Hospitals 454'41	Doors, sound insulation of 451'75:334'5	68-0415

This project has been referred to three A-Z classes, O3, D75 and H41, of which C3 is considered as primary. This is indicated by underlining. To the right we have a current registration number, here fictious. We have the following structure:

A-Z class	C3	AZ-code
1st number	334'5	1st US-concept
1st key-word	Sound insulation of	
2nd number	451'75:454'41	2nd US-concept
2nd key-word or group of words	Doors for hospitals	

Note that the first key-word can act as a hierarchical division within the first number;1) then comes the second concept (number and key-word) not as a hierarchical division but as a relationship; the principle does not work in a predetermined order, it is flexible.

This use of key-words for detailed classification implies that one can be content with relatively short numbers. To go too deep in setting the number can actually prove detrimental, since one then risks the fragmenting of key-words that belong together.

When the key-word lines for all documents are filed together in the order given above we get as index an follows:

O3 Acoustics

....

Sound insulation	Measuring	65-1386
334'5	13	
Sound insulation of	Doors, for hospitals	68-0415
334'5	451'75:454'41	
Sound insulation of	Doors, influence of fissures	66-7413
334'5	451'75:409	
Sound insulation of	Windows, for city dwelling	68-4130
334'5	451'76:454'25	

....

---

1) 451'75 = Doors, gates, shutters, etc.

This method may seem complicated, but in fact it permits more rapid classification than does classification by ordinary methods. One can work consistently, and one has a chance to make an adequate characterization of documents. It is the hesitation between different alternatives, none of which quite seems to fit, and all the locking-up with the help of an alphabetical register, that make ordinary classification for instance with the UDC, both laborious and time-consuming, and the feeling that one yet does get a good solution, makes the work disinclining.

End of quotation from the CIB report.

D. Conformity between adapted systems and the reference system

A certain conformity can exist between the adapted (special) systems and the reference system. If we presuppose that the special systems contain facets or concept series designed by thorough studies, these facets can often be used on lower levels in the reference system. As an example we can take the facets for building elements (D) and building types (H) in AZ. These can be found in US in the appropriate places. Thus "doors" in AZ is D 75, and in US 451'75.

We also can have conformity between different special systems. A special system for traffic may constitute a class in a special system for community planning, and this system may in turn be identical with the class Community Planning in AZ. However, it is probably not possible to get such an accordance over the whole scale.

If an agreement can be reached, the reference system should be standardized, nationally or internationally, for a certain period (e.g. 10 years), while the special systems should be more individually treated, perhaps regulated by societies or organisations representing sectors in society. US is a scetch for a reference system; the best structure for such a system of the future, however, should be object of thorough studies under the guidance of international institutioons.

E. Classification of occupations

The problem of fitting together the employers desire of qualifications for a certain appointment with the qualification of an applicant seems to be closely related to the general documentation problem, where a person's need of knowledge for a special occupation should be matched with the knowledge contained in the stored documents. An appropriate classification is important with regard to the efficiency of the employment exchange.

If we examine the "Nordic (Scandinavian) Classification of Occupations" (N.Y.K), that is based on the International Standard Classification of Occupations, ISCO, we find that the occupations are characterized by the same kind of concepts as in a general classification. E.g.

> House, town, drawing, work, control, work, studies, planning, car, boat, measuring, mines, maps, chemistry, physics, laboratory, geology, genetics, fishes, drugs, agriculture, gardening, wood, hunting, hospital, medical care, school, music, textiles, etcetera.

These concepts are reappearing in many connections. The analogy with product classification is rather close. It may very well be possible to separate a special coarse classification of occupations, representing the general character of occupations, from the specified objects, e.g.

Teacher	Mathematic
	History
	Building
Consultant	Medical
	Juridical
	Building
Labourer	Wood work
	Forestry
	Building

Obvious the problem is more difficult here compared with products, where the industrial processes give a good guidance.

As an example we can take occupations for sawing in wood.We have in N.Y.K.

77 Wood Work
772.10 Frame sawer
.15 Circular log sawer
.20 Cross cutting sawer

These occupations could be expressed as

Labourer - Wood : Sawing (or sawing : wood)

and the analogy with product classification can be illustrated with

Machine - Wood : Sawing (or sawing : wood).

F. A geometrical representation

In fig. 2 the reference system (RS) with its two-dimensional structure is illustrated by a square, where the reference system lies in both horisontal and vertical direction. A combination of two concepts as Wood : Sawing of Sawing : Wood represents two different points in the square.
Along the circle we have the adapted systems, all making use of the same reference system. To the left along the circle we find systems adapted for special trades or branches, building, traffic etc. To the right we have adapted systems of "general type", that is systems which are complete in their kind, adapted not for a special branch but for a certain general category of objects, e.g. products or occupations. May be patents can be included here.

Also a universal document system is indicated. We here have an interesting problem which only can be indicated. In the present debate on the possibility to get a new universal system, there are som people whe advocates that a structure based on traditional disciplines should be maintained, while other experts prefer a universal system based on concepts perhape with a faceted structure.

A discipline system in combination with a concept system is perhaps a solution that should be investigated. On the other side there are situations where one feel a need only of one single system, e.g. by a specialized scientific documentation.

As mentioned above the combination wood : sawing can correspond in product classification to a machine, in occupational classification to a worker or in a universal classification to a heading Material treatmen in the technological class.

The combination pig : liver corresponds to the category Food in the product classification, to Medicine or Zoology in other connections.

G. Example

The table II, page 22 shows codes for documents, products and occupations from three international systems, all with a certain official position: UDC, the Bruxelles Nomenclature and the Scandinavian Classification of Occupations, could be replaced by codes from the AR-complex Here the US-code for pig 636'75 is used sole or together with an A-code for documents, products, products and occupations.

The table shows also how the numbers in many connections can be writte in shorter notations.

Every item can here be introduced in a predetermined position: elephan hides and elephant keepers have their determined positions in analogy with pig hides and pig keepers.

H. Some general aspects on a reference system.

A few words ought to be said about the structure of a future RS. It is probably not possible to state objectively that such a reference system should have this or that structure. Many soultions seen to be possible depending on the designer's own background and interests. US is deduce as a concept system to be used separately for document classification with concept combination. It comes closer to a thesaurus than UDC.

The tests performed so far, however indicate that this system can fill the function of a reference system fairly well by giving rather short notations and to a certain degree one fixed place for every concept. However, as US is originally designed to act independent of another system it has been given a certain resemblance to present universal systems and has connection with our traditional disciplines as an anchorage in something existing and familiar. The separation of the functions, by applying the AR-idea, will perhaps make claims on alterations of the structure in a more semantically determined direction, more independent of technical and social branches.

I. Summary

The advantages of the AR-model can be summarized as follows:

1) Many different quite independent systems can be substituted by a complex of connected systems.

2) Every item (concept) can be entered in a predetermined place.

3) As the codes generally consists of two or more independent numbers the codes can be changed by re-sorting the numbers. The system is flexible.

4) The model can be applied in all areas and classification for bibliographical purposes and documentation can be brought together with classification for statistics and other administrative functions, editing of textbooks etc. We will get a common basic philosophy for classifications of all kinds.

CONNECTION BETWEEN DIFFERENT A-SYSTEMS AND BETWEEN AN A-SYSTEM AND THE R-SYSTEMS

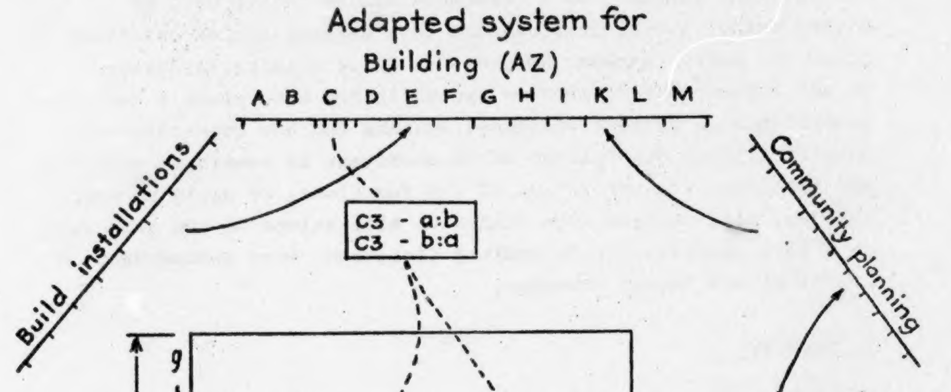
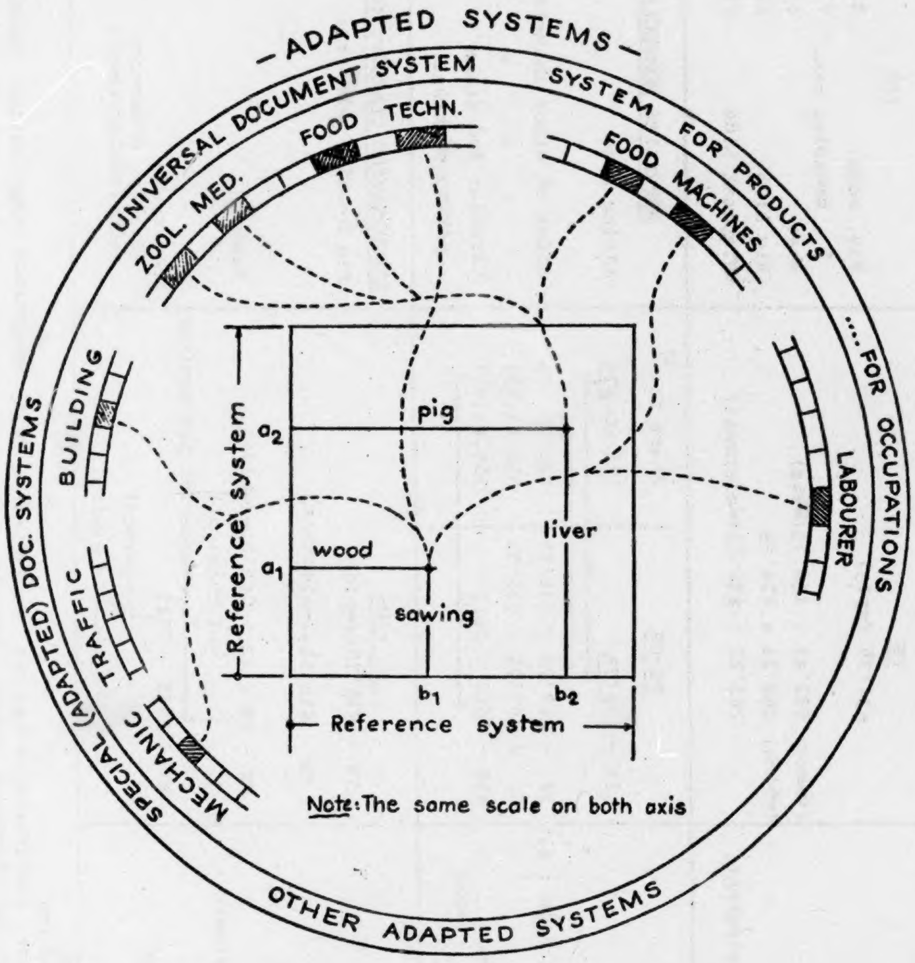

Adapted system for
Building (AZ)
A B C D E F G H I K L M
Build installations
Community planning
Traffic
C3 - a:b
C3 - b:a
Reference system
9
b
a
0
doors
b:a
insulation
sound
insulation
a:b
sound
doors
0 →
a
b
→ 9
Reference system
US.
3
33
334
.1
Accoustics
4
45
451
.7
.75
Building
D in AZ
= D 75
8
86
865
866
Comm. planning
Traffic
a:b = b:a

Fig 1 A A-system (e.g.Traffic) may constitute a class in another A-system (e.g. Community planning) and this latter A-system may in turn constitute a class in a third A-system (e.g.Building). An A-system may also correspond to a subdivision on lower level in the R-system. - All A-systems make use of the same R-system.


- ADAPTED SYSTEMS -
UNIVERSAL DOCUMENT SYSTEM
SYSTEM FOR PRODUCTS
....FOR OCCUPATIONS
SPECIAL (ADAPTED) DOC. SYSTEMS
OTHER ADAPTED SYSTEMS
ZOOL. MED.
FOOD TECHN.
FOOD MACHINES
LABOURER
BUILDING
TRAFFIC
MECHANIC
pig
a_2
a_1
wood
liver
sawing
b_1
b_2
Reference system
Reference system
Note: The same scale on both axis

Firg. 2. Different adapted systems, in the circle, makes use of the same reference system, the square. Within the reference system two-dimensional classification is applied.

TABLE II. COMPARISON BETWEEN THE AR-COMPLEX AND OTHER SYSTEMS

DOCUMENTS	US		UDC	
Pig	636'75 (=pig)		Pig, zool.	599'731'1
			" , breeding etc.	636'4
Pig meat (meat : pig)	(=meat) 682'41 : 636'75(=pig)		Meat	637'525
Pig hide (hide : pig)	(=hide) 263'21 : 636'75		Pig hide	675'03'4 1)
Elephnat hide (hide : elephant)	263'21 : 636'51(=elephant)		Elephnat hide	675'03 : 636'4 2)
PRODUCTS	PC-US	Short No. 3)	BRUXELLES NOMENCLATURE	
Living pigs Living animals - pigs	121 - 636'75	121.00.675	Living pige	01.03
Pig hide Animal struct. prod. - hide : pig	32 - 263'21 : 636'75	320.21.675	Hides of other animals	41.01.903
Elephant hide	32 - 263'21 : 636'51	320.21.651	" " " "	" " "
Machine for sawing in wood Machine tool - sawing : wood	555 - 2862 : 2611	555.62.611	Circular and frame sawing machines	84.47.201/.202
OCCUPATIONS	.... - US		SCANDINAVIAN CLASSIF. OF OCCUPATIONS	
Pig breeder (-keeper) worker - pig	XX - 636'75←(=pig)		Pig breeder (-keeper)	412.40
Elephant keeper worker - elephant	XX - 636'51←(=elephant)			
Tanner worker - material treatment : chemical : hide	XX - 28 : 23 : 263'21←(=hide) (chemical) (=material processing)		Tanner	854.30
Wood sawer worker - sawing : wood	XX - 2862 : 2611 (=wood) (= sawing)		Wood sawer (frame-, cross-cut- and circular sawer)	772.10/.20

1) German edition of UDC
2) Trilingual edition of UDC
3) Reservation: Of course, consideration has not been payed to every imaginable complication. Consequently one must be prepared that the shortened numbers sometimes will be longer than in these examples.

APPENDIX

EXAMPLES ON CLASSIFICATION ACCORDING TO PC - US.

Complete number PC-US	Kind of product	Shortened number	Brussel number
1 NATURAL MATERIALS			
11 Vegetable natural materials			
111 Living plants and parts of plants for reproduction			
111-622'00 .621'5761*)	Apple tree	111.00.5761	06.02
-622'37 .621'3713	Tulip bulb	111.37.3713	06.01
-622'50 .621'3713	Tulip flower	111.50.3713	06.03
-622'70 .621'5323	Seeds of common beet	111.70.5323	12.03.440
112 Plant materials for food (see also 41)			
112-622'25 .621'5323	Common beet	112.25.5323	07.01.219
-622'37 .621'37153	Garlic	112.37.37153	07.01.310
-622'40 .621'8031	Green sallad	112.40.8031	07.01.850
-622'70 .621'33131	Maize grains	112.70.33131	10.05
	- - - - - - - - - -		
12 Animal natural materials			
121 Living animals			
121-636'65	Horses	121.665	01.01.100
122 Fishes and lower animals (see also 41)			
122-631'856	Crayfish	122.1856	03.03.200
-633'513	Baltic herring	122.3513	03.01.153
123 Meat etc. of mammals and birds (see also 41)			
123-6521 Meat			
123-6521 .635'138	Turkeys	123.21.5138	02.02.800
-6521 .636'671	Pigs	123.21.6671	02.01.411
	--------------------		
16 Natural chemicals (Minerals)			
16 -23'31	Borates	16.31	25.30
16 -23'41	Carbonates	16.41	
.2324	Strontium carbonate	16.41.24	25.32.000
16 -23'42	Silicates	16.42	
.2322	Magnesium silicate	16.42.22	
.01	Talc	16.42.22.01	25.27
.05	Asbestos	16.42.22.05	25.24.000
16 -23'27	Chlorides	16.72	
.3312	Sodium chloride (salt)	16.72.13	25.01

REFERENCES

Reports and articles of the author

1. Wåhlin E. "Basic principles and outline of a new universal system of classification".
Conference in Building Classification, Geneva 1949.

2. Wåhlin E. "Principles for a universal system of classification based on certain fundamental concepts and an outline of a variant adapted to technology"
Journal of Documentation, London 1963:4

3. Wåhlin W. "Classification systems and other subjects"
American Documentation, 1966:Oct.

4. Wåhlin E. "Universal Classification"
The National Council of Building Research, Stockholm. Report 1967.

5. Wåhlin E. "Klassificeringssystemens roll i samhället"(The role of classification systems in society)
Teknisk Tidskrift. Stockholm 1967:36

6. Wåhlin E. "Dokumentationens dilemma" (The dilemma of documentation)Teknisk Tidskrift, Stockholm 1968:28

7. Wåhlin E. "Klassificering av materialteknisk litteratur" (Classification of documents in material technology) Report to SINFDOK, Stockholm Jan. 1969

8. Wåhlin E. "Flexibel varuklassificering" (Elexible classification of products /commodities)
Report to SINFDOK, Stockholm Febr. 1969

9. Wåhlin E. "The structuring of information on research projects" CIB, W 52 Conference Dec. 1969

10. Wåhlin E. "Kunskapssystematik och klassifikation, försummade vetenskaper" (Classification and systematics of knowledge, neglected scineces)

11. Wåhlin E. "Ett sarordnat komplex av klassificeringssystem"
(A coordinated complex of classification systems)
Teknisk Tidskrift, Stockholm 1970,No 18

Other documents of current interests

12. Arntz, Helmut "Die DK - eine Vielfacetten - Klassifikation"
Nachrichten für Dokumentation, 1970:4

13. Dahlberg, Ingetraut "Möglichkeiten einer Neugestaltung der DK"
Nachrichten für Dokumentation, 1970:4

14. von Weizsäcker, Carl Friedrich "Die Zukunft der Wissenschaft",
Zeit 1970:50

15. de Grolier, Eric "Quelques travaux récents en matiere de classification encyklopedique"
Bulletin de Biblioteques de Frances 1970,No 3

16. Brown, Harrison "Toward an effective world science information system"
Unesco/ICSU,Sept. 1970

17. Arntz. Helmut "UNISIST in der Realisierungsstufe. Nachrichten für Dokumentation 1970:1

18. Wellisch. Hang". Organisatorische Neuordnun der DK-systems. Nachr. für Dok. 1971:2

19. Classification in science information. A comparative study undertaken by ASLIB for ICSU as a contribution to the ICSU /UNESCO study of the feasability of a World Science Information System (UNISIST),by Brian C. Vickery, June 1969.

A CONCORDANCE
BETWEEN UDC AND THESAURUS OF ENGINEERING AND SCIENTIFIC TERMS /TEST/. RESULTS OF A PILOT PROJECT

Hans WELLISH
University of Maryland,College Park,USA

A pilot project test was undertaken in order to identify the problems involved in using UDC as a switching language for translation of TEST into other languages, and to explore the ways in which such a concordance could be beneficial for the updating and revision of UDC schedules and TEST descriptors. The first part of the test was performed on a sample of terms taken from each of the 177 subject category groups in TEST and thus representative of its scope. Three of these subject groups were also classed in depth; the results of this test will be reported in a later progress report. Statistical analysis of the TEST descriptors and their UDC codes explores the relationships between single and multiple codes, pre- and post-coordinated descriptors; terminological questions and general problems of thesaurus convertibility are also examined and recommendations on the continuatic of the project on a larger scale are given.

Subject retrieval systems in libraries and information centers have always used one or the other of two possible methods: (i) the alphabetical approach, indicating the subject content of documents by index terms taken from natural language and in most cases formalized in a controlled vocabulary known as a subject heading list or more recently as a thesaurus; (ii) the systematic or classified approach, indicating subjects by symbols (other than words) and arranging these symbols in hierarchical order, when an alphabetical index to the systematic order must be used as a verbal access key to the system.

For a long time, discussion has raged back and forth as to the relative merits of the two systems, much heat yet little light being generated in the process. The protagonists of each approach declared from time to time that only their way of doing things would lead the seekers of information to ultimate bliss while the methods employed by the other side would be the surest way to confusion and damnation. In the past, at least one attempt was made to reconcile the two approaches to subject retrieval when the so-called alphabetico-classified system was invented, but since it managed only to combine the least useful features of both systems, it fell soon into well-deserved disrepute. It is only recently that serious and unbiased efforts have been made to find out what basic principles are common to both retrieval methods and to try to arrive at a workable synthesis that would take advantage of the most useful features of both systems.

J. Toman (1) has pointed out that most of the basic principles are common to the alphabetical and the classified approach to subject retrieval, and that the differences between them tend to disappear where multi-dimensional records (such as peek-a-boo cards or machine punch cards) are used. More important still, he showed that there is now a clearly discernible trend to merge the features of alphabetical and classified systems in unified retrieval systems, and that principles that were previously thought to be contradictory are now merged in a new synthesis.

While Toman's insights are of a general nature and not geared to any particular retrieval system, two important new departures in retrieval system design were made in 1970. One is J. Aitchison's "Thesaurofacet" (2) which combines multi-hierarchical classification schedules with an alphabetical index constructed on the thesaurus principle: each term appears both in the classified arrangement where it forms part of a hierarchical system, and in the index where the thesaurus listing shows alternative hierarchies and other relationships not displayed by the classified arrangement.

Another retrieval tool combining the features of a thesaurus and the use of a general classification scheme, compiled by the Swiss Eidgenössische Technische Hochschule (3) is a thesaurus for the fields of electrical engineering, environmental sciences and general concepts where thesaurus terms in German, French and English are linked by corresponding Universal Decimal Classification (UDC) codes in a systematic index.

In line with these developments, the FID's Central Classification Committee decided at its 1970 fall meeting to test the feasibility of a "concordance" between the UDC and the Engineers' Joint Council's Thesaurus of Engineering and Scientific Terms (TEST) (4) which is probably the most comprehensive general thesaurus compiled so far. Since it is known that several countries contemplate a translation of TEST, it was thought to be of great importance to investigate the possibilities of UDC as a "switching language" that would not only alleviate the task of translation but would also serve to generate lists of hierarchically and conceptually related terms on which the usefulness of any thesaurus rests. The compilation of a UDC/TEST concordance, listing one (or more) UDC codes against each TEST descriptor could conceivably achieve the following objectives:

(a) to create a common basis for the compilation of similar thesauri in any of the languages for which suitable UDC schedules exist, by using UDC codes as "switching points" or "nodes";

(b) to test the capacity of UDC to produce codes equivalent in meaning to TEST descriptors;

(c) to fill lacunae in UDC, discovered by critical evaluation of the results of objective (b);

(d) to augment TEST by adding synonyms for concepts found in UDC, either in the form of permitted terms (descriptors) or as non-permitted terms ("use" references);

(e) to compare the categories of Broader terms (BT) and Narrower terms (NT) with the hierarchical arrangement of concepts in UDC with the aim to achieve greater consistency in TEST as well as discovering gaps and/or deficiencies in the hierarchical structure of UDC;

(f) to compare the category of Related terms (RT) in TEST with the cf. ("see") notes in the UDC schedules as well as with the terms used in the alphabetical index; this would also result in a mutually beneficial augmentation of lists of conceptually related descriptors in TEST and of necessary cross-references and index terms in UDC.

The pilot project was undertaken by three experts with many years' experience in the allocation of UDC codes to scientific and technical subjects: Mr. Einar Öhman, Mr. I. de Vries and Mr. H. Wellisch who also served as coordinator of the team and rapporteur to the CCC. Despite the difficulties generated by the fact that the collaborators could not meet personally to discuss methods and problems that arose in the course of the work, since all communications were by way of correspondence only between Sweden, Israel and the U.S., the project proceeded according to schedule. The present progress report summarizes the results achieved during the first phase of the pilot project which was mainly devoted to objective (b). These results, however, have an importance far beyond that of merely compiling a list* of UDC codes for TEST descriptors, which is a relatively trivial task. They attempt to evaluate the UDC as a retrieval tool by comparing it with a modern indexing tool in an objective manner, relying only on statistical evidence and not on subjective value judgments.

Terminology

Throughout this report, terms will be used in the sense given in the following definitions:

Term	A word or words denoting a concept in natural language.
Descriptor	A permitted term in TEST, consisting of one or more terms, e.g. HABITABILITY, SOCIAL ANTHROPOLOGY, RADIO ACTIVATION ANALYSIS.
Identifier	Terms used to denote geographical locations, projects, equipment, trade names, terms taken from an acknowledged nomenclature, code names and specific names of theories and methods. An identifier is not necessarily part of a thesaurus (although TEST uses some), but may be employed in indexing documents.
Single term	A term denoting one concept (mostly indicated

---

* The full list of 177 descriptors with corresponding UDC codes is part of the 1st progress report submitted to FID/CCC in May 1971. Because of the length of this list, only a specimen page is shown in Figure 1.

	by a noun, e.g. ACCOUNTING, or by a noun modified by an adjective or an adjectival noun or gerundive, e.g. BIOLOGICAL BOMBS, EXPOSURE SUITS, CHOKING AGENTS).
Pre-coordinated	Denoting two or more related concepts by a single UDC code or TEST descriptor combined by the makers of the scheme, e.g. MARINE METEOROLOGY, ROCKET ANTENNAS.
Post-coordinated	Denoting two or more related concepts by a combination of two or more UDC codes or TEST descriptors, e.g. "Dissolved oxygen" = DISSOLVED GASES + OXYGEN. The combination of terms is generally not prescribed by the makers of the scheme* but is made at the indexer's discretion.
UDC number	Any string of digits and symbols permitted in UDC, with the exception of the colon.
UDC code	The representation of a concept by one or more UDC numbers. If a code consists of more than one number, these are joined by a colon.
General auxiliaries	All auxiliaries in tables I(c) to I(k), except those in table I(h); also, the 62-1/-9 auxiliaries**.
Special auxiliaries	All auxiliaries in table I(h) and in table II. (Note: the designation of tables follows the one used in the German DK Handausgabe.)

---

[*] The example used in this definition is, however, one of the relatively few that are so prescribed in TEST.

[**] Because of their general applicability throughout classes 62/69 from which the majority (almost 66%) of all UDC codes in the pilot project sample were taken.

Sampling procedure

The descriptors for the sample to be tested were taken from the Subject category index of TEST, since this is the only part of the thesaurus where only permitted terms are listed and also because it was thought that in this way all subject fields covered by the thesaurus would be represented by one descriptor at least. The latter assumption was later found not to be entirely correct since the groups and sub-groups of the Subject category index are not always representative of more or less homogeneous subjects and a considerable amount of overlap exists both conceptually and in the use of descriptors assigned to any one subject group. This is so because the grouping of terms largely reflects the conceptions of a defense-oriented framework based on the COSATI list of subject categories. However, the sampling method chosen ensures the inclusion of all sub-groups no matter how small (e.g. sub-group 2106 NUCLEAR PROPULSION has only four descriptors) which would probably not have been possible if the main alphabetical sequence of TEST had been chosen as the source of sample terms. In fact, a random sample of the alphabetical part, taken for a different purpose (see below, p.16) where two descriptors were taken from each page, yielding 852 descriptors (or roughly five times the number of descriptors in our sample) did not turn up descriptors for some of the smaller sub-groups.

The headings of each sub-group designated by a group number and a bold-face heading were checked for frequency distribution of initial letters and it was found that this distribution coincided roughly with the general frequency of first letters of all descriptors in TEST which was investigated earlier in a paper by Papier and Lin(5).Therefore,a sample of descriptors taken from each subject group and beginning with the first letter of the heading would be representative not only of the subject fields but also roughly of the alphabetical distribution of descriptors in TEST.

The descriptors were then chosen in the following manner: in each group, the first descriptor beginning with the same letter as the heading was chosen; if that descriptor was identical or nearly identical with the heading, the next descriptor was chosen. This was done in order to avoid the occurrence of too many general terms which would have biased the sample towards descriptors that could easily be coded by simple UDC main numbers.

The 177 descriptors chosen by this sampling method included four (PAIN, SEISMIC ARRAYS, CABLE INSULATION and RADIATION DAMAGE) that belonged to more than one sub-group. These were included in the list of descriptors under their respective group headings but were counted only once in all tabulations. The actual number of different descriptors coded by UDC was therefore 173, or slightly less than 1% of the 17.810 TEST descriptors. Admittedly, this is a very small sample, but neither the time allotted to the completion of the first stage of the pilot project nor the circumstances of collaboration between the three investigators made it possible to tackle a larger sample.

In order to test the possibilities of UDC to code not only randomly chosen descriptors but also a more or less well-defined subject group, three additional groups of terms were chosen, each of which was taken from a complete sub-group in the Subject category index. Each of the three investigators chose a subject field with which he was thoroughly familiar and classed every descriptor listed in that group. The three groups chosen were:

0102 Aeronautics
1112 Wood and paper products
1311 Pumps, filters, pipes, tubing and valves

Although the coding of these groups has been largely completed, time did not permit exact tabulation and analysis.

This will be the subject of our second progress report which we hope to be able to submit to the next CCC meeting. It will then also become possible to draw conclusions about the whole pilot project and about aspects not covered by this report, since a larger sample will be available for evaluation.

UDC editions used

The editions used in the compilation of the UDC codes for the descriptors in the sample were the *English Abridged edition* of 1961, designated in the tables by A, the German *Handausgabe* of 1967, designated by M (= Medium edition) and the German Full edition (5th and 8th international edition), supplemented by the English Full edition (4th international edition), both designated by F. Extensive use was made of *Extensions and Corrections* up to and including v. 7, no. 3 , mainly in order to find the most exact UDC code and to check the English terminology for a term found only in the German M or F editions.

The use of editions in different languages and of different degrees of completeness made it sometimes difficult to determine the precise English UDC heading corresponding to a TEST descriptor. Every effort was made to find the official English version as published in an English-language edition of the UDC or in *Extensions and Corrections*; only in cases where no published English term could be found, the German term was translated as faithfully as possible and given in square brackets in the list. The lack of a complete English full edition or at least an English Medium edition was felt to be a serious hindrance, and a full concordance to TEST would probably have to wait until the completion of one of these editions, in order to assure both accuracy and currency of terminology. The problems of terminology will be dealt with later in this report.

Coding

The 177 descriptors resulting from the sampling of TEST were evenly divided between the three investigators, each taking a group of 59 terms. No restrictions regarding the use of different UDC editions were made, the only aim being to assign a UDC code identical to or as nearly equivalent as possible to the TEST descriptors. It was always kept in mind that a full-scale UDC concordance to TEST (or for that matter to any thesaurus) should make it possible for the compiler of a similar thesaurus in another language to use UDC as a "switching language", i.e. find an exact or nearly equivalent term in his own language by looking up the UDC schedule and taking the term or terms found there for incorporation in the target thesaurus. Although this is a highly simplified description of what would have to be done in actual thesaurus "translation", it nevertheless serves to stress the need for expressing TEST terms by identical or synonymous UDC codes. Because of the highly specific nature of many descriptors it was anticipated that heavy use would have to be made of the full editions of UDC, and that the Abridged and possibly also the Medium edition would not be adequate but for the most general terms. The results of the tests did not bear out such an assumption, as can be seen from Table 1. The Abridged edition was fully adequate for 61.9%, the Medium edition had to be used for 18.7% of all descriptors, and the Full edition was needed only for 19.4%, even when the most specific code was assigned to any descriptor. The A and M editions together were thus fully adequate for 80.6% of all descriptors.

It should be noted that where a code had to be taken from M or F, this was necessary in only 20.5% of all cases in order to furnish the main number, whereas it was necessary for only 7.5% of the descriptors in the sample to resort to these editions because of an auxiliary number being given there in full that was shown only partly or not at all in A. An example for the former case would be 0813 SANDS• 624.131.212, where A has only 624.131.2 CLASSIFICATION OF SOILS (no specific soils listed), while an example for the latter case would be 2105 JET ENGINE INLETS 621.454-225.12, where the main number is given in A and only the auxiliary had to be taken from F.

Edition	Most specific UDC code		M or F editions needed for Main number		Auxiliary only	
	N	%	N	%	N	%
A	156	61.9				
M	47	18.7	24	9.5	10	3.9
A + M	203	80.6				
F	49	19.4	28	11.1	9	3.6
Total	252	100.0	52	20.6	19	7.5

Table 1. Editions used for most specific UDC codes.

The picture changes even more in favour of A and/or M editions if the requirement for exact and unambiguous verbal rendering of UDC codes is somewhat relaxed to permit the use of UDC codes whose headings include the equivalent of the TEST descriptor but show also some synonymous terms that are less exact or are only related to the descriptor. An example will make this clear: the descriptor 0616 PAIN has as its exact equivalent UDC 612.884 taken from M where only the word PAIN is given as a heading. If we use UDC 612.88 only which in A has the heading TOUCH. FEEL. TACTILE SENSE. PAIN we would also have an equivalent for PAIN, although the heading is rather broad and includes also related subjects, expressed by other TEST descriptors. (Incidentally, the index to A gives also the full code for PAIN so that in this case A could be said to contain even the most specific code.) If UDC codes

---

• All four-digit numbers preceding descriptors denote the number of their sub-group in TEST.

are accepted as equivalents for TEST descriptors when their headings contain at least one synonym for the concept (and in many instances this would be sufficient for the compiler of a target-language thesaurus since it must be assumed that he would have to have an adequate knowledge of the source language) then the percentages change as shown in Table 2, where A + M are sufficient for 85.3% of all descriptors and only 14.7% have to be taken from F editions.

Edition	UDC codes containing at least one synonym	
	N	%
A	181	71.8
M	34	13.5
A + M	215	85.3
F	37	14.7
Total	252	100.0

Table 2. Editions used for synonymous UDC codes.

Since TEST is now widely used for the indexing of highly sophisticated scientific and technical papers, the results of this analysis would indicate that Medium editions (on the model of the German DK Handausgabe) would be sufficient to fill the needs of classifiers and indexers in a large majority of cases and that the minute detail of full editions would be necessary only for classing material in very specialized fields. This has, of course, always been assumed to be so, but no conclusive evidence has ever been produced to support this assumption. The fact that M and F editions had to be used sometimes only to yield an auxiliary number indicates that a much larger proportion of both general and special auxiliaries should be included in abridged and medium editions since these devices would make exact classing and indexing possible for a majority of concepts denoted by main numbers contained in those editions.

Use of main numbers and auxiliaries

The TEST descriptors were coded as precisely as possible and no restrictions were put on the use of auxiliaries in order to construct the most exact equivalent in UDC. The results show that well over two thirds of the descriptors could be expressed by main numbers alone, even if the most specific terms were employed. The use of auxiliaries was necessary for 31.3% of all UDC numbers, with a fairly even distribution between general and special auxiliaries (see Table 3).

UDC numbers consisting of	N	%
Simple main numbers*	173	68.7
Main number + general auxiliary	34	13.5
Main number + special auxiliary	43	17.0
Main number + verbal extension (A/Z)	2	0.8
Total	252	100.0

* including 4 comprehensive numbers, e.g. 633/639

Table 3. Use of main numbers and auxiliaries.

Use of single and coloned codes

The 173 TEST descriptors yielded 212 UDC codes, the majority of which (81.6%) were expressed by single numbers (simple main numbers or main numbers specified by auxiliaries). Only relatively few descriptors had to be expressed by coloned codes (18.4%), all but one of which were composed of two numbers (one coloned code consisted of three numbers). There was only one term that had no UDC equivalent, namely 1202 QUEUING

THEORY, although even in that case a number has been proposed several years ago but has not yet been accepted officially because of an on-going large scale revision of class 51 Mathematics.

Kind of code used	N	%
Single	173	81.6
Coloned	39	18.4
Total	212	100.0

Table 4. Use of single and coloned codes.

Use of one or more alternative codes

The UDC is essentially a multi-aspect classification, while a thesaurus is generally believed to be more oriented towards unique assignment of terms to concepts (except for the inevitable occurrence of homonyms which are taken care of by suitable indication of the context). It was therefore surprising to find that the pilot project resulted in a large number of unique UDC equivalents for TEST descriptors and that multiple choices occurred relatively seldom. It seems that the UDC after all does not have as many multiple notations for the same concept as is sometimes assumed to be the case. Even in one instance that needed no less than five different UDC codes, the fault lies rather with TEST which lists 0603 BAMBOO only once under Biology, whereas UDC has codes for it in Botany, Agriculture and in three different fields of application; only the three latter ones could be said to be redundant and might in fact in the near future be eliminated by the adoption of a general Materials facet in UDC. Table 5 shows the distribution of single and multiple codes in the sample.

TEST descriptors		UDC codes		
N	%	Number of alternat.	Number of codes	%
139	80.3	1	139	65.6
28	16.2	2	56	26.4
4	2.3	3	12	5.7
1	0.6	5	5	2.3
172	100.0		212	100.0

Table 5. Use of alternative codes.

Regarding multiple use of the same descriptor in TEST, the results are even more surprising. Among the 173 different descriptors in the sample, 59 (34.1%) belong to two different TEST subject groups and 4 descriptors (2.3%) belong to three different subject groups. (As indicated before, 4 of the two-group descriptors appear in the sample itself as duplicates.) Since it was thought that these high percentages might not be representative because of the small size of the sample, a larger sample was tested for this phenomenon by counting the occurrence of descriptors belonging to two or three different subject groups in two randomly chosen sequences of 10 pages each. Pages 226-235 yielded 450 descriptors and pages 373-382 yielded 396 descriptors, together 846 descriptors or 4.7% of all TEST descriptors. It was found that 113 descriptors (13.3%) belonged to two subject groups and only one descriptor (0.1%) belonged to three. Taking these percentages as representative, about 2400 TEST descriptors would belong to two subject groups and about 180 descriptors would appear in three groups. This result shows that the often deplored tendency in UDC to offer "multiple choices" in coding the same concept in different categories is also prevalent in TEST and probably in many other thesauri.

Pre-coordination versus post-coordination

The substitution of UDC codes for TEST descriptors is equivalent to translating from one language into another, or more specifically from one indexing language to another. Although, as has been shown above, the concepts covered by TEST can be expressed in practically all cases by a suitable UDC code, there are considerable differences between the two indexing languages as far as their syntactical structure is concerned.

TEST uses almost exclusively single terms or pre- coordinated terms as descriptors, but one or more of the constituent parts of a pre-coordinated term are in many instances also listed separately as single terms. This is almost the rule when a single term is used to denote a genus whose species are denoted by pre-coordinated terms where the generic term often becomes the second part. This leads to the well-known dispersal of closely-related concepts throughout the alphabet. An example may be the term RATIONS which occurs also in the more specific pre-coordinated terms EMERGENCY RATIONS, FIELD RATIONS, FLIGHT RATIONS, JUNGLE RATIONS, MILITARY RATIONS, etc. The indexer, to be sure, will find these and other kinds of Rations as NT in the TEST tables (even though the choice between FIELD RATIONS and MILITARY RATIONS might not always be made consistently). In retrieval, on the other hand, this must inevitably lead to recall failures because a searcher entering a file under FIELD RATIONS is not aware of the fact that there is also closely related material under MILITARY RATIONS nor is it indicated to him that general material on the subject will be found under RATIONS, unless documents have been indexed consistently under both single (generic) terms and by their permitted combinations (specific terms).

UDC, on the other hand, is a largely synthetic indexing language that relies on post-coordination of concepts, either in the form of a main number specified by an auxiliary or in the form of two (or more) coloned main numbers, denoting the elements of a complex concept, yet making it possible to retrieve the constituent parts also independently. In either case, the main concept can be found in one place, further modified by an auxiliary or a coloned main number. The above example of RATIONS would be classed in UDC by 355.653 RATIONS, the context of 355.6 MILITARY ADMINISTRATION making it clear that these are MILITARY or FIELD RATIONS (which takes care of the near-synonymity of the two TEST terms). Next, 355.653 : 355.469 might be FLIGHT RATIONS (assuming that these rations are used during air operations) and 355.653 (253) would be JUNGLE RATIONS. In any case, all information on RATIONS will be found together, while related information on JUNGLES and FLIGHTS can also be retrieved (independently or in connection with RATIONS).

Pre-coordinated codes occur, of course, also in UDC, such as for example 368.44 UNEMPLOYMENT INSURANCE which is simply an enumerative extension of 368.4 WORKMEN'S INSURANCE despite the fact that 331.6 UNEMPLOYMENT appears as a subdivision of 331 LABOUR and could have been coloned to 368.4. Such pre-coordinated codes for complex concepts are now considered to be an unfortunate part of the inheritance from the original Dewey Decimal tables and systematic efforts are being made to avoid such pre-coordination in new schedules or to eliminate existing pre-coordinated codes by suitable revision.

Despite such differences in syntax it was found that the vocabulary of the two indexing languages could be matched in the overwhelming majority of cases, if exact equivalents as well as synonyms are accepted as matches. Less than 10% of all TEST descriptors in the sample could not be expressed by suitable UDC codes which is much less than had been expected, especially since the terminology of at least some English UDC editions is now somewhat out of date.

The decision as to whether a TEST descriptor was to be counted as a single or pre-coordinated term descriptor was necessarily a subjective one, based on the following criteria: a descriptor was counted as "single term" when it consisted of

a) a single word, e.g. AERONOMY, SANDS, CANNING, or
b) a single concept expressed by two or more words, e.g. FIRST AID (but not FIRST AID KITS), CHARACTER RECOGNITION, FOOT AND MOUTH DISEASE*

There were many borderline cases such as CAVITATION CORROSION which were mostly counted as "single terms", interpreting this designation rather liberally.

Applying these criteria to the 173 descriptors in the sample, it was found that 109 descriptors (63.1%) fell into the category of "Single term descriptor" and 64 (36.9%) were "pre-coordinated descriptors". Post-coordinated terms, although permitted in a few instances in TEST**, could not appear in the sample since they are not included as possible descriptors in the Subject category index.

In order to discover whether this distribution of descriptors is representative of the whole TEST, a representative sample of all descriptors in the Alphabetical list was taken. Two terms per page were checked, taking the last descriptor in the lower left-hand column and the first descriptor in the upper right-hand column. If any of these descriptors was a non-permitted term, the next higher (or lower) permitted term was taken, except that non-permitted terms with an instruction to use post-coordinated descriptors were also counted, so as to find out the use of this mode of term construction (of which the introduction says on p. 8 that it "occurs relatively few times"). The alphabetical listing of descriptors occupies 426 pages which yielded 852 terms, or a sample of 4.7% of all descriptors. Table 6 shows the distribution of descriptors in the whole TEST as compared with the distribution of descriptors in the pilot project sample (PPS).

---

* Some of these terms are taken from the full TEST list and do not form part of the sample.

** e. g. DISSOLVED OXYGEN is to be expressed by DISSOLVED GASES + OXYGEN.

Descriptors	TEST		PPS	
	N	%	N	%
Single term	572	67.1	109	63.1
Pre-coordinated terms	276	32.4	64	36.9
Post-coordinated terms	4	0.5	-	-
Total	852	100.0	173	100.0

Table 6. Distribution of 3 kinds of descriptors in 2 samples from TEST.

A comparison of the two samples shows close agreement between the percentage figures; very roughly, two thirds of all descriptors in TEST as well as in the pilot project sample are single term descriptors, the remaining third being pre-coordinated descriptors (the number of post-coordinated terms is insignificant). This result also tends to reinforce the hypothesis that the pilot project sample of descriptors is truly representative of the character of all TEST descriptors, despite the rather small size of the sample.

Since it was believed that the occurrence of identifiers in TEST might influence the applicability of UDC codes to such TEST descriptors, a count of these was also made in the sample taken from the full TEST. There is no distinction between Descriptors (for concepts) and Identifiers (for specifically named entities) but as stated in the introduction to TEST (p. 4) certain categories of terms have been excluded from it. Nevertheless, some TEST descriptors are essentially identifiers, such as the trade names of certain products (e. g. BAKELITE, NUVISTORS), the names of theories or equations (e. g. GINSBURG-LANDAU THEORY, MAXWELL'S EQUATION), names of diseases (e. g. GAUCHER'S DISEASE, TURNER'S SYNDROME) or codes and acronyms (e. g. HMX for Cyclotetramethylenetetranitramine). It was found that 19 or 3.3% of the 572 "single term" descriptors in the tested sample were such identifiers which, by definition, are single term descriptors; they would, however, not always have to be expressed by a UDC number (except by adding a verbal extension A/Z to a main number, e. g. 678.6 BAK for Bakelite). If these identifiers are deducted from the number of "single term" descriptors, however, the proportions are not significantly changed, as shown in table 7.

Descriptors	N	%
Single term (minus identifiers)	553	66.4
Pre-coordinated terms	276	33.1
Post-coordinated terms	4	0.4
Total	833	100.0

Table 7. Data as in Table 6 less Identifiers.

Terminology

It has often been claimed that the terminology of TEST and other thesauri is much more precise than anything that could be found in the UDC tables, but this assumption has been shown to be quite unfounded. As can be seen from table 8, no less than 31.6% or almost a third of the TEST descriptors had *exact* equivalents in UDC headings, while another 58.9% were adequately expressed by synonyms. Using the UDC as a switching language for the construction of thesauri in languages other than English, a very high percentage of all descrip ors (possibly as high as 90%, as in the pilot project sample) could easily be found in the UDC schedules either as identical or as synonymous terms. Since all thesauri contain synonyms of descriptors as non-permitted terms with attached "use" references, the list of such synonyms would only have to be augmented by those found in UDC headings, to make the thesaurus both more compatible with UDC and more useful to users who would be given additional entry words to the thesaurus vocabulary. As mentioned above, the compiler of a foreign-language target thesaurus would have to be familiar enough with the source language to be able to decide on the best equivalent term given among those listed in the UDC schedule.

In order to compare the terminology of TEST and UDC by a method that precluded subjective judgments as far as possible, the following procedure was used: when a UDC code chosen to express a TEST descriptor had a heading that consisted of the same word or words it was counted as an exact equivalent. When the UDC heading consisted of several terms, one of which was an exact equivalent, only that term was listed and counted. If there was no exact equivalent term in the heading, the nearest synonym only was chosen for the list. All other terms in the heading were disregarded as being either redundant or not suitable. For example: the TEST descriptor 0203 AGRICULTURAL MACHINERY was coded by UDC 631.3 Agricultural implements, tools, machinery. From this heading, only the words AGRICULTURAL ... MACHINERY were taken as an exact equivalent. The TEST descriptor 0814 GEOELECTRICITY was coded by UDC 550.37 Terrestrial electricity. Earth currents, from which only the first term was taken as the most nearly synonymous one.

In many instances, a term listed in a specific UDC heading may be too general or ambiguous if taken only by itself; such terms must always be read in conjunction with the containing heading (at the next higher hierarchical level). Mainly for this reason, the generic term to any specific UDC term has always been indicated in the list.

Table 8 shows the extent to which TEST descriptors could be expressed by exactly equivalent terms or by synonymous terms and which descriptors could be subsumed under broader terms only, i. e. no specific equivalent could be found. As already noted above, only one term could not be expressed at all by a UDC code. In the following analysis of table 8 examples for each category of term substitution are given, and those terms that involved terminological problems are dealt with more thoroughly. The first term in each example is taken from TEST, followed by the UDC code and its heading as given in the list.

Single term descriptors (STD)

For 36 STD in this category exact equivalents expressed by a single code were found, e.g.

ACCOUNTING - 657 - ACCOUNTING
OCTANES - 547.217.2 - OCTANES

42 STD were expressed by a single code and one or more synonyms, e.g.

EARNINGS - 331.2 - WAGES. SALARIES. PAY
GEODESICS - 513.573 - GEODESIC LINES*

15 STD were expressed by a post-coordinate code, consisting of a main number and an auxiliary, yielding an exact verbal equivalent, e.g.

CACHEXIA - 616-056.55 - CACHEXIA
FABRICS - 677.06 - FABRICS

16 STD were also rendered by the same kind of post-coordinate code, but had synonyms as headings, e.g.

CHARACTER RECOGNITION - 681.3.071.2 - COMPUTERS-DIRECT READING DEVICES
GAS GUNS - 623.42.083.4 - GUNS-GAS FILLINGS

---

* Incidentally, this example shows how UDC terminology could benefit from a check against TEST terminology: there are three alternatives for GEODESICS in UDC, two of which have the heading GEODESIC LINES and one has the heading GEODETIC LINES.

12 STD were expressed by coloned codes and synonymous headings, e.g.

MARINAS - 627.21: - PORTS (ACCORDING TO PURPOSE):
797.1 - BOATING

Pre-coordinated term descriptors (PTD)

5 PTD had a pre-coordinated single UDC code and an exact verbal equivalent, e.g.

INDUSTRIAL HYGIENE - 613.6 - INDUSTRIAL HYGIENE
MAGNESIUM ALLOYS - 669.721.5 - MAGNESIUM ALLOYS

12 PTD had pre-coordinated single UDC codes with a synonymous heading, e.g.

INDUSTRIAL HYGIENE - 628.5 - INDUSTRIAL SANITATION
CARTOGRAPHIC CAMERAS - 528.711.11 - PHOTOGRAMMETRIC CAMERAS
or 771.319.5 " "

In these cases, UDC is clearly in need of revision, since the pre-coordinated codes violate the rule to express complex concepts by codes that allow for the independent retrieval of each component. This can be demonstrated in the case of CARTOGRAPHIC CAMERAS which should be expressed by two coloned codes, one for PHOTOGRAMMETRIC TECHNIQUES AND APPARATUS and one for CAMERAS (SPECIALIZED, FOR SCIENTIFIC PURPOSES).

11 PTD were expressed by a post-coordinate code in the form of a main number and an auxiliary, yielding an exact verbal equivalent, e.g.

ROCKET RESEARCH - 629.76.001.5 - ROCKETS-RESEARCH
INDIUM ISOTOPES - 546.682.02 - INDIUM-ISOTOPES

21 PTD were also rendered by a main number and an auxiliary, but had synonymous headings, e.g.

CABLE TOOL BITS - 622.243.414.051 - CABLE DRILLING-BITS
JET ENGINE INLETS - 621.454-225.12 - JET ENGINES-INTAKES

22 PTD were expressed by coloned codes and synonymous headings, e. g.

WOUND BALLISTICS - 531.58: - TARGET BALLISTICS:
616.001.45 WOUNDS CAUSED BY FIREARMS

MISSILE FUZES - 623.451.8: - ROCKET MISSILES:
623.454.2 PROJECTILE FUZES

As noted above, all these account for more than 90% of all descriptors in the sample.

We must now turn our attention to the remaining 10% that could not be expressed adequately and specifically by a UDC code and its heading, but only by subsuming them under broader terms or by constructing a code that somehow approximated the meaning of the TEST descriptor.

6 STD could be expressed by a single code but not with the exact meaning of the descriptor; such TEST descriptors should be added to the UDC schedules and in most cases a new code would have to found. An example for such a case is

ASTROGEOLOGY - 523.05 - ASTROGEOLOGY

This number was constructed by analogy, since 523.03 is Astrophysics, 523.04 Astrochemistry and 523.07 Astrobiology, where the last digit corresponds to the last digit of the main number in class 5, viz. 53 Physics, 54 Chemistry and 57 Biology. It would seem permissible to construct a number 523.05 Astrogeolog taking the final digit 5 from 55 Geology*. This is an instance where TEST terminology clearly points to a gap in UDC that should be filled as soon as possible. Some other examples are

PINCH VALVES - 621.646.221 - SIMPLE VALVES
SAFE HANDLING - 658.382.3 - SAFETY MEASURES-WORK MANAGEMENT

6 STD could be expressed by a main number and an auxiliary, but no equivalent term was found to give the exact meaning of the TEST descriptors, e.g.

SENSORY DEPRIVATION - 616.89-008.42 - SENSORY DISORDERS-ILLUSIONS, ETC.

---

* This does not imply that we consider the use of .0 subdivisions as main numbers as a permissible notational device. Quite to the contrary, we believe that this usage should be eradicated from the UDC tables wherever it appears. But as long as such numbers are officially sanctioned, we must use them, however reluctantly.

(The concept "Sensory deprivation" is here coded as a pathological phenomenon; see below for another attempt to code this concept.)

MEDIUM RANGE BALLISTIC MISSILES - 623.451.8(1-194.4) -
ROCKET MISSILES-MIDDLE DISTANCE

This is certainly a clumsy way to denote "Medium range". It would be better to develop a special auxiliary for 623.451 Projectiles. Bombs, to denote the range (e.g. short, medium, long, inter-continental).

5 STD could be denoted by a coloned code that approximated the meaning of the TEST descriptor, e.g.

HABITABILITY - 577.4: - GENERAL CONDITIONS OF LIFE (IN) RELATION TO ENVIRONMENT:
599.9 - MAN IN NATURE

The term "Habitability" would preferably have to be expressed by a UDC code in one of the envisaged subdivisions of the new class 4, as one of the concepts related to the topic of Environment, ecology, etc. Another example would be

SENSORY DEPRIVATION - 159.3: - SENSE PERCEPTION:
159.97 ABNORMAL PSYCHOLOGY

Here, the concept has been treated in the context of Psychology (following TEST usage), but the rendering is entirely inadequate, as are indeed the entire UDC schedules for psychology (now in the process of revision).

In the category of TEST descriptors lacking UDC codes but which should be included in UDC, only one term, QUEUING THEORY, was found. As already indicated above, a number has in the meantime been proposed and will no doubt become part of the UDC with the publication of new schedules for 51 Mathematics.

There is also a category of TEST descriptors that are lacking a UDC equivalent but need not be included. The sample did not contain such terms, but it is known that TEST contains a number of descriptors that are used by military authorities in the United States only and would therefore not be needed in any thesaurus written in the language of another country (and

TEST descriptors			UDC codes										
	N	%	Single codes (1) (2)		Post-coordinated codes (5)			No specific equivalent (4)			Lacking in UDC (3a) (3c, 10)		All UDC codes
			Exact equivalent	Synonym	Exact equivalent	Synonym		Single code	Post-coordinate code		To be added	Not wanted	
					M + A	M + A	M : M		M + A	M : M			
Single term	109	63.1	36	42	15	16	12	6	6	5	1	0	
Pre-coordinated term	64	36.9	5	12	11	21	22	2	1				
Sub-total	173	100.0	41	54	26	37	34	8	7	5	1	0	
Total	173		95		97			20			1	0	212
Percent			44.8		45.8			9.4					100.0

M + A = Main number and auxiliary
M : M = Coloned main numbers

Table 8. Translation of TEST terms into UDC codes.

.393

not even in an English-language thesaurus used in the U.K. or in Australia, etc.). Examples of such terms are

CLASS 1 SUPPLIES
GA AGENT

and similar local expressions. The lack of an equivalent UDC code would in these cases have beneficial effects and would act as a kind of filter for terms that are not needed in the target-language thesaurus, although they are part of the source thesaurus.

The numbers in parentheses given in the column headings of table 8 refer to an article by H.H. Neville (6) in which he demonstrates 11 different cases that might arise when converting the vocabulary of one thesaurus to that of another one in the same or related fields. Although there are some differences between the "reconciling of terms" in two thesauri and the conversion of the terms of one thesaurus to codes in a classification scheme, the similarities between these two operations are more important and, in any case, the findings of Neville are relevant to our topic.

Neville's cases 1, 2, 3a, 3c, 4, 5 and 10 are exactly equivalent to the ones covered by the categories shown in table 8. The remaining cases are:

Case 6, where A (the source thesaurus) distinguishes between homonyms, while B (the target thesaurus) does not. UDC has no need to distinguish between homonyms (except in the alphabetical index) because the hierarchy of terms always makes the context clear: 556.36 is SPRINGS and so is 621.888.4, but the subordination of the former to Earth sciences and the latter to Workshop techniques distinguishes between the two meanings of the term SPRINGS.

Case 7, where A uses separate descriptors to distinguish between different uses of a term, while B does not distinguish between them. Again, UDC generally solves this problem by using the same term in different hierarchies, e.g. (to take

Neville's own example) ALUMINIUM and ALUMINIUM STRUCTURES appear in 669 Metallurgy and in 69 Construction industry respectively. TEST has numerous such descriptors and most if not all of them would be taken care of by the use of codes taken from appropriate hierarchies in UDC.

Case 8, where A uses descriptors of a general nature such as METHODS, TYPES, HIGH, LOW, etc. that are not found in B. This problem is largely solved in UDC by the use of common auxiliaries, especially the so-called Point-of-view numbers, which can be used to denote METHODS, TYPES, DESIGN, EQUIPMENT, etc. TEST includes relatively few of these general descriptors and their use is generally not recommended (scope notes appended to such descriptors make this clear). General adjectival descriptors of the type HIGH, LOW do not occur in TEST at all.

Case 9, where A contains synonymous descriptors that are used in different contexts. Also this case is resolved by the use of codes taken from different hierarchies in UDC. Neville actually cites an example from TEST: the descriptor FREQUENCIES is used in connexion with "cycle times" and "waves", but not with "tones" for which PITCH is prescribed. However, PITCH is also used for the bituminous material. The corresponding UDC codes would, of course, be taken from 534 Accustics and from 665 Oils, Fats. Waxes, respectively.

Neville's article has been treated here at some length because he recommends the use of arbitrary numbers as links between the terms of the source thesaurus and those of the target thesaurus. Yet there is, of course, no reason why UDC codes could and should not be used in this capacity. This would indeed be highly advantageous to the work of "reconciliation" by displaying hierarchical groupings of related terms, revealing terminological deficiencies in both thesauri and pointing out related terms that had not been included in either one.

Distribution of UDC classes in the TEST sample

Table 9 shows the distribution of 2-digit classes represented by UDC numbers in the list. Since TEST, as its title implies, is mainly devoted to science and technology, the bulk of numbers (about 91%) are naturally distributed throughout classes 5 and 6, a pattern familiar to most UDC users.

In class 5, 53 Physics is most heavily used, followed by 55 Earth sciences, 57 Biology and 54 Chemistry. Class 56 Paleontology is not represented at all, but this subject field has only a few descriptors in TEST, so that its absence in the sample is to be expected.

In class 6, 62 Engineering and technology is, of course, the largest group, followed at some distance by 61 Medicine, 65 Management and 66 Chemical technology.

Class 3 comes next in frequency of use, mainly because of class 35 or rather 355 Military science which furnishes all 6 numbers used. In a heavily defense-oriented thesaurus this is only to be expected. The placement of Military science among the social sciences has always been deplored as one of the worst mistakes in the hierarchical structure of UDC and is one of the unfortunate inheritances from Dewey.

Classes 2 and 9 are not used at all. A check of the relevant groups in TEST, viz. 0504 History, law and political science and 0506 Humanities, shows that in the former group, only the descriptor HISTORY itself would qualify for a number in class 9, while in the latter group there are only two descriptors that would fall in class 2, namely CHRISTIANITY and RELIGION. Again, in our small sample, no representation could be expected.

	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	
	0		1		2		3		5		6		7		8		9		
	00	2													80	1			
	01	1							51	4	61	17							
	02	1							52	6	62	101							
							33	3	53	23	63	9							
							34	1	54	7	64	1							
	05	1	15	2			35	6	55	10	65	1[illegible]							
			16	1							66	11							
									57	9	67	7	77	2					
									58	3	68	6							
							39	1	59	1	69	1	79	1					
Total		5		3		0		11		63		166		3		1		0	252
%		1.9		1.2		0.0		4.4		25.0		65.9		1.2		0.4		0.0	100.0

C = Class

N = Number

Table 9. Use of UDC numbers by classes.

Conclusions and recommendations

(1) The results of the pilot project show that it is possible to find corresponding UDC codes for a very large percentage (possibly up to 90%) of all descriptors in TEST without sacrificing exactness in the meaning of terms. UDC is thus highly suitable as a "switching language" used for the translation of TEST (or similar thesauri of a general or discipline-oriented nature) into other languages and every effort should be made to undertake this task on a full scale. It would largely be a one-time project the costs of which would be recovered in the saving of untold hours of individual efforts in thesaurus compilation; it would also avoid costly mistakes and trial-and-error procedures now adopted in the construction of foreign-

language thesauri. It would have to be updated from time to time, possibly with every new edition of the source thesaurus, but the bulk of UDC codes would remain stable since they denote concepts independent of any natural language and are therefore not affected by mere changes in terminology.

(2) More than 80% of all TEST descriptors could be covered by UDC codes taken from Medium editions. The efforts to publish new UDC schedules in various languages should therefore be concentrated on these editions rather than on the perfection of Full tables since the latter are probably most used by subject specialists who would in any case have to know sufficiently much English or German to be able to use the Full schedules already published in these languages. On the other hand, the UDC auxiliaries should be given in full (or almost so) in the Medium editions since combinations of main numbers (to the extent given in Medium editions) plus Auxiliaries allow for the coding of most descriptors.

(3) The terminology of UDC should be brought up to date wherever a comparison with corresponding descriptors in TEST shows that new terms are now used for those that are found in UDC headings for the same concept. Care should be taken, however, not to confuse Americanisms with new terminology, and appropriate British, Canadian, Australian, etc. terms should also be included in TEST.

(4) Synonyns for a concept as given in UDC headings, if not found in TEST either as descriptors or as "use" references, should be incorporated in TEST as additional parts of the lead-in vocabulary.

(5) There are concepts in TEST not covered anywhere in UDC (or only in a roundabout and clumsy way). These should be given preferential treatment in the extension and revision procedures of UDC and both new codes and/or suitable additions to existing ones should be made.

800.92UDC

THE UDC
AS LANGUAGE OF INFORMATION IN ITSELF

Jan H. de WIJN
Netherlands Nider, Hague, Netherland

Introduction

The natural language we speak, is an aid for human beings to formulate their thoughts and to convey their ideas to others. Unfortunately one can never be sure that others - the receivers of the message - actually understand what he - the sender - intends to say.

If I ask you:

> "And climate do girls Herceg-Novi how in like the the you?"

I have an idea that you do not understand what I am asking you, even if you may have heard some well-known sounds like: "girls" and "Herceg-Novi". The sentence I have spoken however contains all the necessary elements. If I reword it, it runs:

> "Do you like the girls and how the climate in Herceg-Novi?"

Now the sentence contains two questions:

- "Do you like the girls?"
- "How (...) the climate in Herceg-Novi?"

The second part is incomplete, because we do not know, whether the question is:

- "How is the climate?, or
- "How do youllike the climate?"

But when I ask you:

> "How do you like the girls and the climate in Herceg-Novi?",

you will certainly understand that I have asked:

> "How do you like the girls (in Herceg-Novi) and (how do you like) the climate in Herceg-Novi?".

Without going too deeply into details, it will be clear that a message to be carried from one to another not only has to contain all the informatory elements, but that there has also to be an agreement on how to use them.

The language we speak has to be formulated in such a way that its message will be onderstood by all the people for which it is intended. For this purpose there has to be an agreement between speaker and listener on the meaning, both are to give to terms and combinations of sounds.

If the contact between speaker and listener or between sender and receiver does not take place by sounds or by conversation, but in writing, further agreements have to be made on the notation of these sounds.

A specific Dutch example is that they refer to the "Crazy Englishmen", because they:

- write: street (pronounced as straight)
- say: striet
- mean: straat (the Dutch word for street)

One of the most universal notations for sounds is the music-notation, understood all over the world.

Rules and conventions

Language agreements have grown up into language rules and conventions, but unfortunately they differ from language to language or for groups of languages. Nevertheless they are fundamental conditions for the exact use of those languages.

As we have learned from the example, not only the notation of sounds, but also the sequence of the elements and, generally speaking, the whole mock-up or construction of the sentence is subject of agreement. Slight differences in the sequence of the words may give a sentence a totally different meaning.
In a German textbook, I found under a drawing of a violin-player:

pupil: "Die Macht der Musik" (the power of music)
teacher: "Der macht die Musik" (he makes music)

The same letters and the same words, but another sequence, resulting in a totally different meaning.

A similar problem arises with the addition of restrictions or limitations of the meaning. In my own language, like in several other languages, it is usual that restrictions of the meaning of substantives are indicated by adjectives, generally placed before the substantive, or otherwise by specific endings, for instance:

- hond (dog)	witte hond (white dog)
- vriend (friend)	vriendin (girl friend)

In the English language for instance, endings are less frequently used, but in the Russian language it is quite normal to indicate differences by means of endings:

Ivan Borissovitsj	Ivan, son of Boris
Olga Sergejevna	Olga, daughter of Sergej

Artificial languages

In the same way, the practical use of artificial languages used in the field of information - the transfer of knowledge - is bound to specific rules and conventions, based on agreements.

In information-languages, that make use of elements - words or terms - taken from a natural language, these elements usually suggest a certain meaning. In fact this is more an image than a reality, because such a word may have several meanings.
In th Dutch language for instance, the word "school" has the following meanings:

- institute for education (teaching)
- the programme of such an institute
- the total of persons taught together
- the school-building
- a group of proselytes
- a group of the same animals, e.g.: herrings
- a separate part of a fishing-net

You all know that especially in specific fields a word may have a special meaning. In an information language therefore an agreement must be made on the use of a specific term to represent a specific and simple conception. This is the origin or source of descriptor-lists and thesauri.

In information-languages, that do not use words taken from a natural language, there is no relation between the "notation" and the specific conception. It is almost a matter of dispute, whether this lack of relation is an advantage or a disadvantage, but it is a fact that if we come to an agreement on the use of a specific way of spelling or notation, it does not matter, whether this is done by words or by other notational elements, such as the UDC.
What method you prefer depends on other factors, which I shall leave out of discussion for the moment.

Each "notation", maybe composed of words or of other means, consists of two distinct sorts of elements:

- primary elements, to be used as search-characteristics
- secondary elements, to function as subdivision-characteristics

The first category supplies the approach to the system and the second category limits the field-of-search.
In contrast to some natural languages, the secondary elements in information-languages - and especially in the UDC - are placed after the primary elements, independent the mode of spelling or notation:

- J.H. de Wijn	in bibliographies:	Wijn, J.H. de
- White dog		Dog, white
- Stroomtransformator		Transformator, stroom

In systems using symbols, the same practice is followed:

- H	Information retrieval
Hg	Notation
Hgf	Faceted notation

- B	Verpackungsmittel
B8	Kunststoffpackungen
B82	Flaschen aus Kunststoff
B8299	Spezialflaschen aus Kunststoff
B8291	Sprühflaschen aus Kunststoff

- 4	Materials
44	Plastics
441	Plastics by form and behaviour
4413	Foam-expanded plastics

- 1	Filosofie
17	Zedeleer. Levenswijsheid
178	Ethiek en onthouding
178.7	Tabaksverslaafdheid

From these examples, taken from different classifications, we learn that they all follow the same course and in respect hereto, there is also less difference.

The Universal Decimal Classification

According to the foregoing, the UDC is an artificial information or indexing language, that uses figures and symbols and by means of these elements a decimal notation is built up.
Wellisch formulated the decimal principle and fundamental law for the UDC as:

> "Any element in a complex UDC-code specifies the element written at its left side up to, but not including, any relational sign that may precede that element."

The UDC as an information language has its own rules and conventions, these have to be learned and especially their application has to be practised. It is a dangerous fallacy that one can learn and use the UDC without knowledge of the "Rules of the game" and without any training.
If I have in my hand a booklet containing Yugoslavian words with their Dutch equivalents, I shall not yet be able to "speak" the Yugoslavian language. Even my English is still far from faultless.

One of the great advantages of the UDC is that a very complex conception can be expressed by simple expedients, because the UDC diposes of a whole series of elements, numbers and symbols of different levels.
In the same way as in natural languages the adjective limits a substantive, the meaning of a UDC number can be restricted within certain limits by the addition of auxiliaries. In the example "Fancy dog" for instance, the conception "dog" is restricted to nice, pretty, fancy dogs only and all ugly dogs are excluded.
In the UDC this is expressed by adding a next decimal digit:

636.7	Honden	Dogs	Sobaki
.76	Schoothonden	Fancy, pretty dogs	Konatnye sobaki

The main purpose of the addition of attributes to a conception is to limit the field of search and it is important to limit such a field with the aid of as few elements as possible, but also as far and as speedily as possible.
In the UDC this can be done by adding those elements to the number representing the core-subject, that limit the meaning of that number most, for instance:

001.891	Research
-05	Research-workers
-055.2	Female research-workers

Theoretically, the total number of research-workers is limited to one half if there are as many female as male workers. If we consider all the research-workers and we envisage only those in the field of documentation, there are perhaps only a few thousand on the world, and only in the Netherlands, there are only some hundreds available.

001.891-05::002	Research-workers in the field of documentation
(492)	---- in the Netherlands

If we are aware that the UDC-number has to be used to classify documents according to the subject given, e.g. reports in the Russian language, we shall find a very small number of documents or their surrogates in our file, indicated by the notation:

001.891-055.2::002(492)(047.3)=82

Limitations by attributes

The expressions attributes, facetes, auxiliaries or relations may refer to several matters.
Vickery for instance, in the track of Ranganathan, gives the conception "facet" a very specific meaning:

Vickery:	Ranganathan:
Whole thing	Personality
kinds	
parts	
materials	Matter
properties	
processes	Energy
operations	
agents	
space	Space
time	Time

The UDC does not distinguish those "facets" in this way, but groups the attributes according to their limiting function, their frequancy of use or to their field or level of application.
Generally speaking, the UDC main number always indicates the "entity", the "whole thing" or the "personality". The numbers are grouped according to characteristics in common, so that we see that not only auxiliaries are used to indicate characteristics, for instance:

621.313.3	alternating-current electric machines
.8	magneto-electric machines

are grouped together into:

621.313	electric machines
.31	electric power supply
.3	electrical engineering

but subdivided into:

621.313.32	synchronous machines
.33	asynchronous machines
.36	commutator (alternating-current electric) machines

The facets "space" and "time" however, together with "form" and "language" (of the document) and "race or nationality" are dealt with as separate facets in the UDC also.

Austin, one of the promotors of the faceted-classification and member of the British Classification Research Group, starts also from the "thing" or "entity", and recognises the following "relations":

-thing-to-kind or type
-thing-to-parts or elements
-thing-to-attributes
-thing-to-other thing

Compared with the UDC, Austin considers kind or type as well as parts and elements as the "main-numbers", forming the natural hierarchy.
Since "parts and elements" are considered as "things" also, the relation "thing-to-parts" is considered as a "thing-to-other thing" relation.
In the UDC we find then the same construction:

things	main numbers
attributes	auxiliary numbers
relations	colon

As Austin, at least theoretically, does not group his "things", he obtains a very broad and shallow classification that can only be used for specific fields. In practice however, "grouping" is used in faceted-classification also as shown for instance in the "Classification of Library Science" as used for the periodical: "Library and Information Science Abstracts":

E/F	Organization and Administration
Ec	Buildings
Ef	Equipment
Fr	Staff

See also appendix IV

Rules of the Game

The purpose of the UDC, like the purpose of all indexing languages, is to make documents or data retrievable, and the notation serves to limit the field of search. A logical consistency is that in indexing documents as well as in the "translation" of inquiries or queries into UDC-numbers, the same methodology - the same rules and conventions - for the building up of the notation has to be followed.

Fundamentally the UDC has the disposal of three groups of elements:

I	Main numbers
II	Auxiliary numbers
III	Connection signs

An inventory of all UDC-elements is given in appendix I.

The main-numbers have to give the translation of the conception or core-subject to be indexed, but sometimes one simple main-number may be not enough and a combination of two or more numbers has to be used to represent one single conception, for instance:

54+66	Chemistry and Chemical Industry
629+656	Transport and Traffic
(435.9+492+493)	Benelux
669.2/.8	Non-ferro metals being: 669.2+669.3+.....+669.8
669.15'24'26	Chromium-nickel-steel alloy being: 669.15+669.24+669.26

The auxiliary-numbers have to limit the field and they are available in several categories. There is a tendency to subdivide them into:

- special auxiliaries, with a more or the less limited field of application and another meaning in different fields
- common auxiliaries, applicable to the entire UDC with almost the same meaning
- artificial auxiliaries, not scheduled in the UDC, but formed by the aid of:

a. main UDC numbers	::...
b. non-UDC codes	=...
c. words	A/Z
d. non-UDC numbers	No...

No exact order for the application of the auxiliaries is prescribed, but in practice, those elements are taken first that greatly limit the conception or subject and next those elements that are of interest, but less important for retrieval or limitation of the field of search.

Compared with the bibliographical description of a document, the author and the subject are of primary interest for retrieval.
The place and date of issue of the document may be of secondary importance, but although the colour of the binding or the number of pages gives us some information, these features will never be used for retrieval and we seldom find a collection of red books or of papers consisting of 93 pages.

The connection signs are used to combine two or more notations for separate conceptions into a complex notation. In this respect the colon is frequently used and therefore specification of its meaning is proposed for instance by Perreault relators or by Wellisch relation signs.

Document analysis

For the purpose of document analysis, the UDC-elements are to be distinguished also in another way:

- contents analysis or analysis of the information itself, being the translation of contents features into UDC notational elements (subject analysis, "internal" characteristics)
- formal analysis or analysis of the information carrier, being the recognition of bibliographical elements and other "external" cha characteristics and the translation of them into notational elements.
  Within a UDC-notation normally only the document form and the language of the document are given, but it is possible to indicate place and date of issue and other data also.

A special category of elements is that indicating the viewpoint, from which the subject described was handled. It will be clear that this element best can be added after the number for the subject is determined, thus after the combination of elements indicating the subject itself.

When analysing a document, the following scheme can be used:

1.	first subject		next subject	
2.	attributes		attributes	
3.	viewpoint		viewpoint	
4.		relation		
5.				information carrier (form, language, etc.)

e.g. 123-45.006.7 : 987.065.000.2 (047.3)=86

Upon developing this scheme into a model for document analysis, the following "model" was elaborated:

See appendix II.

If a UDC-notation is built up with the aid of this model, it will always be reproducable and can be read like natural languages.
When decoding such a notation, one shall always find the original conception. This model, used as a check-list, assures us that in indexing as well as in retrieval the UDC will effectively function as an information or indexing language

Citation order and filing order

Since the model functions only as a check-list, you will understand that even a sequence is suggested, in fact no citation order is prescribed. The citation order depends only on the analysis of the very document.

The filing of document surrogates, even done by human beings or by machines, has to be done mechanically, and depends in no way on the analysis of the document, its contents or on the interpretation by the document-analyser.
Although the filing order already was fixed a long time ago, Wellisch drew the attention of the CCC to the fact that the filing order recommended in different UDC-editions was not exactly the same and sometimes incomplete.
For that reason, the CCC-Committee "UDC/Fundamentals" recommended a new filing order:

See appendix III.

Though this filing order will only be recommended, it is clear that especially within a specific file, or within a set of collaborating files only one filing order can exist.

Mr. Chairman, Ladies and Gentlemen,

I have endeavoured to show you the relation between natural and artificial languages and I have especially tried to indicate the need for rules and conventions for an effective workability of all those languages, and especially for the usability of the UDC.
I have also shown you that the incoherent talk of the first example cannot be used for information and is in no way a "language".
This is in force for all languages, including the UDC.
The use of the UDC as a language of information or indexing in itself requires stringent rules and also their exact application.
The broader the circle of users of the UDC is, the stronger is the need for those rules and regulations.

I thank you very much for your kind attention

The Hague, 1971-03-25

Ing. J.H. de Wijn

appendixes:

- I Survey of UDC notational elements
- II Model for Document Analysis and Representation by UDC
- III Filing order of UDC notational elements
- IV Outline of "Classification of Library Science"

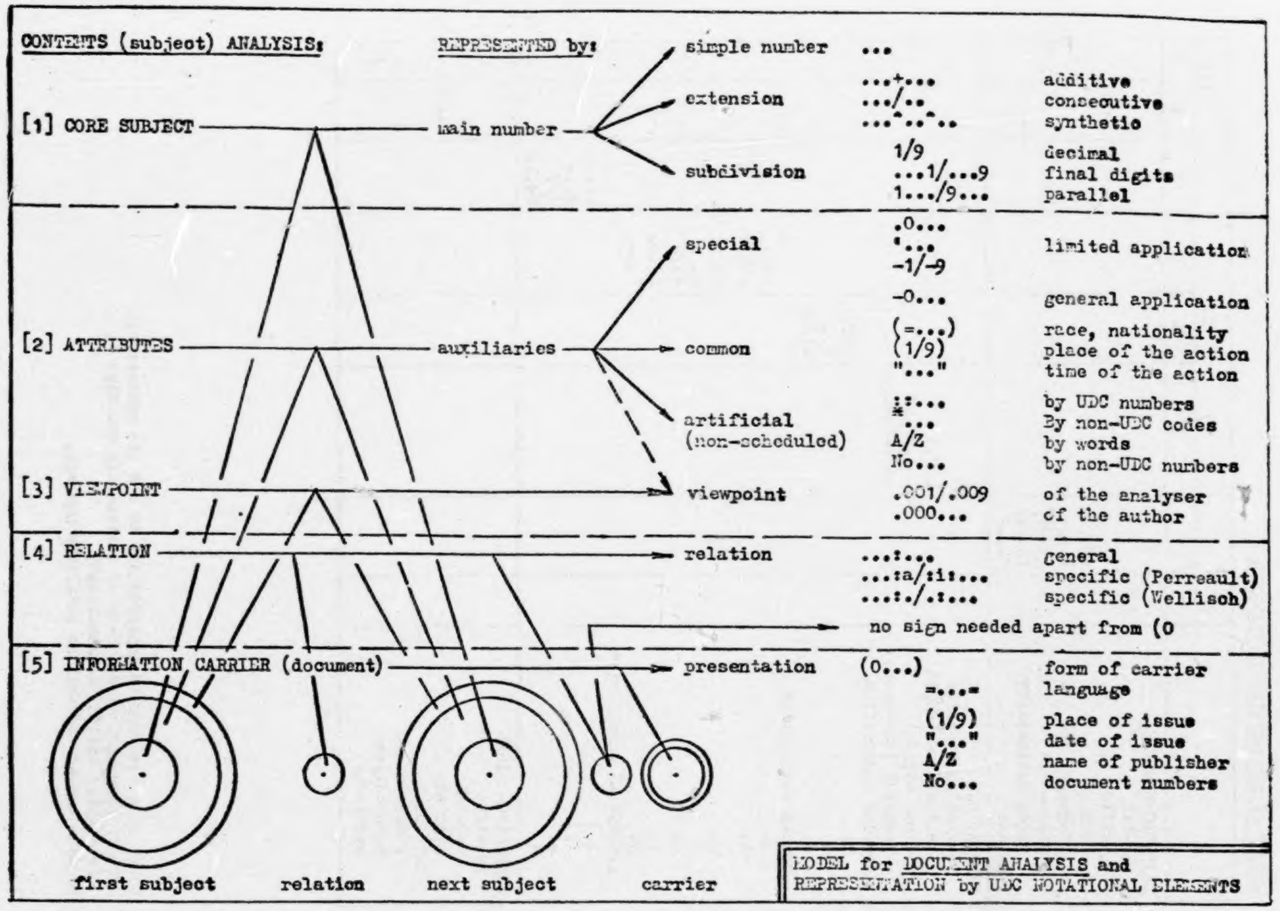
appendix I

SURVEY of UDC NOTATIONAL ELEMENTS

	I		II				III	
	A	B	C	D	E	F	G	H
MAIN UDC NUMBERS								
General	0/9							
Specific								
form		(0...)						
language		=...=						
place		(1/9)						
race, nationality		(=...)						
time		"..."						
LIMITATIONS								
Decimal subdivision								
next decimal digit			.1/.9					
final digits			...1/...9					
parallel (analogous)			.1.../.9...					
Special auxiliaries								
a.				.0...				
b.				'...				
c.				-1/-9				
Common auxiliaries								
a.					-0...			
b.					.00...			
c.					(0...			
d.					=...			
e.					(1/9			
f.					(=...			
g.					"...			
Artificial auxiliaries								
a.						::...		
b.						*...		
c.						A/Z		
d.						No...		
CONNECTION SIGNS								
Relation								
general							:	
special							:a/:i	
Extension								
additive								+
consecutive								/
synthetic								^

There is a tendency to consider and to use the numbers for
form, language, place, race or nationality and time
(sometimes called "detachables")
as main numbers as well as auxiliary numbers.

appendix II


CONTENTS (subject) ANALYSIS:
REPRESENTED by:
[1] CORE SUBJECT
main number
simple number ...
extension
...+... additive
.../.. consecutive
...^..^.. synthetic
subdivision
1/9 decimal
...1/...9 final digits
1.../9... parallel
[2] ATTRIBUTES
auxiliaries
special
.0...
'...
-1/-9
limited application
-0... general application
common
(=...) race, nationality
(1/9) place of the action
"..." time of the action
artificial
(non-scheduled)
:... by UDC numbers
*... By non-UDC codes
A/Z by words
No... by non-UDC numbers
[3] VIEWPOINT
viewpoint
.001/.009 of the analyser
.000... of the author
[4] RELATION
relation
...:... general
...:a/:i:... specific (Perreault)
...:./.:... specific (Wellisch)
no sign needed apart from (0
[5] INFORMATION CARRIER (document)
presentation
(0...) form of carrier
=...= language
(1/9) place of issue
"..." date of issue
A/Z name of publisher
No... document numbers
first subject
relation
next subject
carrier

MODEL for DOCUMENT ANALYSIS and
REPRESENTATION by UDC NOTATIONAL ELEMENTS

FILING ORDER OF UDC NOTATIONAL ELEMENTS

Extension	+ / ^					additive consecutive synthetic (not yet accepted)
Simple main number		0/9				
Relation			: :a/:i			general specific (Perreault relators)
Limitation				=...= (0..) (1/9) (=...) "..."		(independent the use as auxiliary numbers or as specific main numbers)
				::... *... A/Z No...		formed by UDC main numbers non-UDC codes words non-UDC numbers
				.00... -0...		alteration into -00... proposed
				-1/-9 '... .0...		as new special aux. proposed
Subdivision					.1/.9	decimal digits (independent their origin)
Next UDC number		0/9				

appendix IV

CORE SUBJECTS – A/Z

A	LIBRARY SCIENCE
Ab/Ay	Common subdivisions (Form)
B	Common subdivisions (Subject)
Bd	Biographies (Individual)
Bg	Profession
Bgc	Education
C	Common subdivisions (Time) History Sub-divided by BNB Period Subdivisions
D	Common subdivisions (Place) Sub-divided by BNB Locality Subdivisions
E/F	Organization and administration
Ec	Buildings
Ef	Equipment
Fr	Staff
G	Processing of library materials
H/J	Information retrieval
Hf	Classification
Hm	Cataloguing
Ho/Jrw	Subject indexing
Jv	Author cataloguing
K	Information work. Information science
L/M	Library materials and stock
L	Library stock by physical characteristics
M(0-9)	Library materials by subject. Sub-divided by UDC.
N/Z	Library ownership and use
Nm/Py	Extent and nature of service
P	Library services (Kinds of access to material)
Q	Use of library materials
S	Libraries in general
Sd	Special libraries
Sl	Libraries by subject interest. Sub-divided by UDC
T/V	Libraries by owner-cum-user
T	Private and non-official institution libraries
U	Academic libraries
Ul	University libraries
Ux	School libraries
V	Official institution libraries
Vs	Public libraries
W/Z	Libraries by user
Ww	Children's libraries
X	Student use of libraries
Y	Use by the old, handicapped and other social groups
Z	Users with special subject interests, activities or occupations. Sub-divided by UDC.

FRINGE SUBJECTS—1/9

15/18	Organization and management
16	Data processing
21	Knowledge and learning
23	Education
26	Research
28	Museums
29	Archives
3	Authorship
4/45	Reading
46/48	Writing
5/54	Bibliography
55	Paper
61/65	Printing
66/68	Documentary reproduction (Reprography)
69	Bookbinding
7/77	Publishing
78	Bookselling
8	Non-book, non-printed materials
87	Microforms
9	Other subjects. Sub-divided by UDC.

FINAL SESSION

SUMMARY OF SYMPOSIUM CONSENSUS AND POSSIBLE FUTURE TRENDS

Chairman: Mr. G.A. Lloyd, FID Classification Dept.
Members: Mr. P.G. Beak
Mr. D. Fink
Mr. R.M. Hansen
Mrs. M.T. Martinelli
Mrs. I. Wesley
Mrs. B. Živković

Each of the panel members gave 5-10 minute impressions of the papers and discussions at the sessions during the first three days of the Symposium, and these were followed by a general discussion in which several participants joined. The Chairman, Mr. G.A. Lloyd, then attempted to summarize the consensus of the Symposium and the likely trends in the UDC development programme for the seventies.

1) The role of the UDC in the 1970s

FID's central effort should focus on the reference or switching role of UDC in major information systems and networks such as the UNISIST Project of UNESCO and ICSU, and on the necessary improvement of the superstructure rather than on piecemeal detailed revisions which should be the responsibility mainly of FID/C... committees and other special group interests. This does not mean that detailed development and updating of the UDC should slow down: the P-note and "Extensions" services based on the regular work of FID/C... committees are essential for the many thousands of users of existing UDC schedules available in a score or more of different languages and of different scopes - full, medium, abridged and special - subject editions.

It is obvious, however, that a system like the UDC, with its superstructure so akin to the 100-year-old Dewey DC, cannot without major structural changes continue to serve as an internationally accepted ordering system for the rapidly proliferating new subject-fields of knowledge that have to be handled in modern information retrieval systems. Only a drastic modernization of the superstructure can fit the UDC for this role, and it must be the main CCC task to bring about the essential improvement as soon as possible, in such a way as to satisfy the new demands while causing minimum inconvenience to old UDC users the world over.

2) Basis for UDC improvement

UDC improvement should be based on better methodology (theory and principles) and clearer guidelines for those working to revise specific fields. This was the message of several Symposium papers, notably that of A. Neelameghan (India), and even if obliquely-of Ejnar Wåhlin, and of Inge Dahlberg in her plea for an 'ideal' indexing language to replace the UDC. Clearly stated principles and correspondingly more helpful directives are essential at all levels of UDC work during the seventies: they must be set out by the CCC in a better version of UDC Revision Procedure and then applied, as far as possible, as well to the detailed routine revisions by FID/C... committees as to the elaboration of the new superstructure.

3) A new Standard Reference scheme

It would seem to follow from the two foregoing points that the FID itself should aim at preparing, publishing and maintaining a Standard Reference (SR) scheme consisting of a 'roof classification schedule' of not more than three or four hierarchic levels as the unifying/switching language, with suitably indicated points and symbol devices for coupling to specific classifications, thesauri, etc., leaving the main responsibility for full editions with national members and for special-subject editions mainly with the specialized interests concerned, national or international.

If the superstructure of this SR scheme has to be so radically changed that virtually a new Universal Classification emerges, it may be necessary to distinguish the SR notation from the 'orthodox UDC' and to refer or couple to detailed UDC notations in the same way as to other classifications or thesauri.

A really sound superstructure that reflects modern consensus, however, might well gain acceptance - eventually, if not immediately- among wide circles of present UDC users who are (often reluctantly) being forced to contemplate alternative schemes, provided that its notation is so designed as to permit direct coupling with the many quite satisfactory detailed divisions of the existing UDC schedules. If this can be achieved swiftly and relatively painlessly, there is a good chance that the new SR scheme could play the dual role of an international unifying/referral code for UNISIST and similar information systems, and a more detailed artificial subject-indexing language - the New DC- that could be adopted by countless new and old UDC users in multilingual and multidisciplinary information centres throughout the world, both in computer-aided and traditional (manual) modes.

G.A. Lloyd

Concluding speech of Mr. Ljubomir Dulović

At the end of our Symposium my speech will be very short. First of all I would like to point out once again what were our objectives when we gave the proposal to organize the symposium "UDC in Relation to Other Indexing Languages" They were:

1. to discuss from all points of views about the real value of UDC and its relation to other indexing languages
2. to evaluate real possibilities of using UDC in information systems, and
3. to give an opportunity to participants from Yugoslavia to learn about the state-of-the-art of indexing languages and UDC and experiences of representatives from developed countries.

I think we have to be satisfied. About 19 papers and discussion showed that we were right. Symposium proceedings will be valuable documents for many scientists in information science. It is the contribution of all of us to the new science - information science.

It is my duty to apologize where we were not be able to satisfy all wishes concerning accomodation and weather, especially at the end of our Symposium. But we must be also ready to sacrifice something for progress and prosperity of human beings as well as many known and unknown members of our society (like many discoverers or now cosmonauts).

I hope to meet all of you at the Third Seminar "UDC and mechanization" which will be in Yugoslavia in 1973, perhaps also in Herceg Novi.

LIST OF PARTICIPANTS

Foreign participants

1. Arnaud P., France
2. Beak P.G., Commonwealth Agricultural Bureau, Oxford, England
3. Bystrich E., Germany
4. Cooke G.A., Canada
5. Cruzeiro M.M., Portugal
6. Černyj A., USSR
7. Dahlberg I., Germany
8. Diakonova O., UNESCO, Paris
9. Dreese M.J., Holland
10. Fink D., Denmark
11. Gervasi M., Italy
12. Gunton D., U.K.
13. Hindson R., U.K.
14. Lloyd G.A., FID, The Hague
15. Martinelli M.T., FAO, Rome
16. Mølgaard Hansen R., Denmark
17. Neelameghan A., India
18. Öhman E., Sweden

19. Perreault J., USA
20. Rigby M., USA
21. Samuelsson K., Sweden
22. Wählin E., Sweden
23. Wegelius P., Finland
24. de Wijn J.H., Netherlands
25. Lobova Lj., SSSR
26. Thalberg K.R., Normway

Participants from Yugoslavia

1. Aleksić D.
2. Anić Ž.
3. Baskić R.
4. Bavec M.
5. Borisavljević B.
6. Borovnjak J.
7. Bratić I.
8. Brborić B.
9. Čampara E.
10. Čolić R.
11. Debeljuh M.
12. Djaković V.
13. Djenić B.
14. Djurić M.
15. Djurin Lj.
16. Dulović Lj.
17. Filipović D.
18. Gavrilov M.
19. Gigić R.
20. Glumac D.
21. Godec M.
22. Grlić V.
23. Horvatović Z.
24. Ignjatović Ž.
25. Ikonić B.
26. Ivančević Lj.
27. Janković R.
28. Jekovec J.
29. Jerković B.
30. Jeršek V.
31. Ježova P.
32. Josipović M.

33. Jovanović N.
34. Jungwirth N.
35. Karalić E.
36. Kiauta D.
37. Knežević M.
38. Kokole J.
39. Kokot N.
40. Kos S.
41. Kujović O.
42. Kupčević Lj.
43. Lalić R.
44. Lazarević B.
45. Leitner A.
46. Leota A.
47. Lorber F.
48. Lukežić M.
49. Luković V.
50. Madjarić M.
51. Maslovarić A.
52. Mikačić M.
53. Milinčević M.
54. Milovanović S.
55. Mirković M.
56. Mladenović N.
57. Mole M.
58. Mulalić Z.
59. Nikolić Dj.
60. Nonin Lj.
61. Obradović S.
62. Pašalić Lj.
63. Patrnogić M.
64. Pavunić M.
65. Perić M.
66. Perpar G.
67. Peručić M.
68. Pešić D.
69. Petrović N.
70. Polojac Z.
71. Popović E.
72. Proguter A.
73. Prohić E.
74. Radović M.
75. Rešetar S.
76. Sajkov G.
77. Sofronijević M.
78. Spasić D.
79. Stamatović D.
80. Strugar K.
81. Škrk J.
82. Toldi Dj.
83. Topić Š.
84. Tošić V.

85. Trećakov E.
86. Umek M.
87. Vasiljević A.
88. Vićentijević D.
89. Višnjić M.
90. Vujačić M.
91. Vukasović K.
92. Wesley I.
93. Živković B.
94. Žlindra J.